Exact Analytical Loop Closure in Proteins Using Polynomial Equations

WILLIAM J. WEDEMEYER, HAROLD A. SCHERAGA

Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853-1301

Received 23 November 1998; accepted 8 February 1999

ABSTRACT: Loop closure in proteins has been studied actively for over 25 years. Using spherical geometry and polynomial equations, several loop-closure problems in proteins are solved exactly by reducing them to the determination of the real roots of a polynomial. Loops of seven, eight, and nine atoms are treated explicitly, including the tripeptide and disulfide-bonded loop-closure problems. The number of valid loop closures can be evaluated by the method of Sturm chains, which counts the number of real roots of a polynomial. Longer loops can be treated by three methods: by sampling enough dihedral angles to reduce the problem to a soluble loop-closure problem; by applying the loop-closure algorithm hierarchically; or by decimating the chain into independently moving rigid elements that can be reconnected using loop-closure algorithms. Applications of the methods to docking, homology modeling and NMR problems are discussed. © 1999 John Wiley & Sons, Inc. J Comput Chem 20: 819–844, 1999

Keywords: loop closure; ring closure; rigid geometry; spherical geometry; resultants

This article is dedicated with affection to the memory of Ulrich Kinkel.

W. J. Wedemeyer was an NIH Trainee, 1993–1996.

Correspondence to: H. A. Scheraga; e-mail: has5@cornell.edu.

Contract/grant sponsor: National Science Foundation; contract/grant numbers: DMR-9632275 and MCB95-13167.

Contract/grant sponsor: National Institutes of Health; contract/grant number GM-14312.

This article includes Supplementary Material available from the authors upon request or via the Internet at ftp.wiley.com/public/journals/jcc/suppmat/20/819 or http://journals.wiley.com/jcc/

CCC 0192-8651 / 99 / 080819-26

Introduction

Loop closure has been a focus of research from the earliest days of computational protein modeling.[1–3] The n-atom loop-closure problem is defined here as follows: given a chain of n atoms linked by rigid bonds at fixed bond angles, and given the (fixed) positions of the first two and last two atoms, determine all possible positions of the $n-4$ intervening atoms that satisfy the bond lengths and bond angles. This article derives mathematically exact solutions for seven-, eight-, and nine-atom loop-closure problems, by reducing them to finding the real roots of a polynomial. In particular, the (nine-atom) tripeptide and (eight-atom) disulfide loop-closure problems are solved. Extensions of the method to longer loops are also given.

The loop-closure problem received its first analysis in the articles of Gō and Scheraga.[1,2] The first article[1] demonstrated that at least 6 degrees of freedom are required to close a loop of rigid bond lengths and bond angles, because six constraints must be satisfied in the translation and rotation of a coordinate frame from the first residue to the final residue. Gō and Scheraga formulated these constraints as six equations in the loop dihedral angles, and applied these equations to several particular cases, for example, tripeptide loops[1] and ring molecules such as the cyclic decapeptide Gramicidin-S.[4] It was found empirically that as many as eight solutions are possible for the closure of a tripeptide loop,[1] although the reasons for this limit have remained obscure. It should be noted that the Gō–Scheraga equations must be solved numerically, for example, by Newton's method,[5] and that finding multiple solutions requires multiple starts.

The Gō–Scheraga equations do not guarantee that even one solution exists for closing a given loop with a prescribed rigid geometry. For example, the original article showed that cyclic tripeptides and tetrapeptides cannot be closed if the peptide groups are all *trans*.[1] In such cases, additional degrees of freedom must be added to permit loop closure, for example, additional dihedral angles[6–8] or bond angles.[9] Of course, even these extended methods will fail to find a solution if the six translational and rotational constraints are severe enough.[6,9]

The Gō–Scheraga equations are completely general, and may be applied to loops longer than a tripeptide.[1,2,4,6,11] However, their solution becomes more difficult when solving for nonadjacent dihedral angles. An alternative approach is to model the translational and rotational constraints by an energy (or penalty) function, which can be minimized to obtain loop closure to a good approximation.[12–15] Once satisfied, these constraints can be maintained under minimization, Monte Carlo, or molecular dynamics if changes in the dihedral angles are kept small (quasi-infinitesimal) and made orthogonal to the loop-closure constraint forces.[16–18] Other methods of loop closure rely on distance geometry methods[19,20] or on searching a database of loop conformations drawn from high-resolution protein structures.[21–24]

Loop-closure algorithms have found several uses in protein science. For example, they have been applied successfully to modeling proteins from experimental NMR distance constraints[25,26] or from the known structure of a homologous protein,[24,27–30] as well as to docking problems.[31] Loop-closure algorithms are also needed for constructing all-atom models of flexible loops whose electron density may be poorly defined in the crystal structure.

Moreover, as pointed out in the original article,[1] loop-closure algorithms permit local rearrangements, i.e., concerted changes in dihedral angles that rearrange a short segment without disturbing the rest of the protein. In particular, Monte Carlo simulations would benefit from the ability to make local rearrangements efficiently. Unfortunately, local rearrangements are difficult when proteins are modeled by rigid geometry (as in ECEPP[12,32,33]), because each dihedral angle moves all subsequent atoms and, thus, even small changes of a single dihedral angle can result in large atomic motions, with correspondingly large changes in the energy. Local rearrangements are also difficult when proteins are modeled with flexible bond lengths and bond angles, because any conformation which violates the ''normal'' bond lengths and bond angles produces a very high energy, which must subsequently be relaxed by computationally expensive minimizations or Monte Carlo. Thus, an efficient, exact loop-closure algorithm is useful for sampling protein conformations, regardless of whether rigid or flexible geometry is adopted. This article derives such an algorithm using spherical geometry and polynomial equations.

This article is arranged as follows. In the first section, the tripeptide loop-closure problem is introduced, and the geometric variables are defined. In the second section, nine unit vectors are pro-

jected onto the unit sphere, forming three spherical quadrangles. Using results derived in Appendices A–C, these quadrangles give rise to three coupled biquadratic equations in three variables, which are solved in Appendix D, resulting in a single polynomial equation of degree 16 in one variable. The real roots of this polynomial correspond to the valid closures of the tripeptide loop. The third section computes bounds on valid solutions of the three biquadratic equations, and methods for checking whether the loop closure polynomial has any solutions. In particular, the method of Sturm chains allows the number of valid loop closures to be computed. The fourth section shows that the tripeptide loop closure problem is equivalent to the general seven-atom loop-closure problem. The fifth section solves the problem of closing disulfide-bonded loops (an eight-atom loop-closure problem), which results in a polynomial of degree 32. The final section discusses extensions of the method to longer loops.

Throughout this article, the term ''biquadratic'' refers to a polynomial equation in two variables that is of degree 4 overall but of degree 2 in each variable, i.e., a biquadratic equation is quadratic in each variable when the other variable is regarded as a parameter. Similarly, ''biquartic'' refers to a polynomial equation in two variables that is of degree 8 overall but of degree 4 in each variable, i.e., a biquartic equation is quartic in each variable when the other variable is regarded as a parameter.

Geometric Definitions

The tripeptide loop closing problem is depicted in Figure 1. The positions of four atoms (N_1, C_1^α, C_3^α, and C_3') are fixed while the positions of the five intermediate atoms (C_1', N_2, C_2^α, C_2', and N_3) remain to be determined. Six dihedral angles (ϕ_1, ψ_1, ϕ_2, ψ_2, ϕ_3, and ψ_3) are undetermined, although only three dihedral angles are needed to position the five atoms; the remaining three dihedral angles can be determined from the atomic positions once the constraints of loop closure have been satisfied.[18] For example, the dihedral angles ϕ_1 and ψ_1 fixes the positions of C_1', N_2, and C_2^α, while the dihedral angle ψ_3 fixes the positions of C_2' and N_3; the three remaining dihedral angles (ϕ_2, ψ_2, and ϕ_3) are fixed once the positions of these atoms are determined. To keep the presentation simple, the peptide bonds are assumed to be planar, although the tripeptide loop closure problem can be solved with the peptide bond dihedral angles fixed at any values.

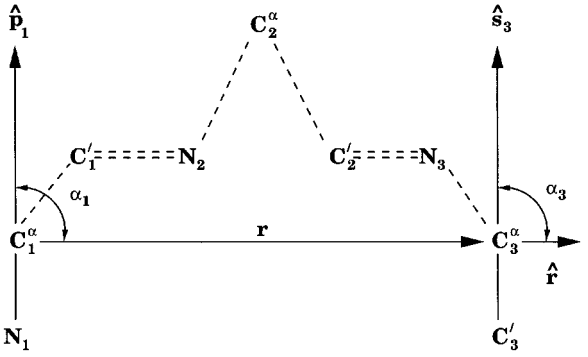

FIGURE 1. The problem of closing a tripeptide loop consists of finding the positions of five atoms (C_1', N_2, C_2^α, C_2', and N_3), which satisfy the bond length and bond angle constraints. The four fixed atoms (N_1, C_1^α, C_3^α, and C_3') define the vector **r**, the unit vectors $\hat{\mathbf{p}}_1$ and $\hat{\mathbf{s}}_3$ and the angles α_1 and α_3.

The positions of the four fixed atoms (N_1, C_1^α, C_3^α, and C_3') determine several essential vectors and angles (Fig. 1). The first is the vector **r** between C_1^α and C_3^α and its corresponding unit vector $\hat{\mathbf{r}}$. Two others are the unit vectors $\hat{\mathbf{p}}_1$ and $\hat{\mathbf{s}}_3$, which point along the N_1—C_1^α and C_3'—C_3^α bonds, respectively. The angles α_1 and α_3 denote the angles between the unit vectors $\hat{\mathbf{p}}_1$ and $\hat{\mathbf{r}}$, and between $\hat{\mathbf{s}}_3$ and $\hat{\mathbf{r}}$, respectively.

Several other unit vectors may also be defined (Figs. 2 and 3). The vectors $\mathbf{q}_1$ and $\mathbf{q}_3$ represent the virtual bond vectors[34] connecting C_1^α—C_2^α and C_3^α—C_2^α, respectively; their unit vectors are denoted by $\hat{\mathbf{q}}_1$ and $\hat{\mathbf{q}}_3$, respectively (Fig. 2). The unit vectors $\hat{\mathbf{s}}_1$, $\hat{\mathbf{p}}_2$, $\hat{\mathbf{s}}_2$, and $\hat{\mathbf{p}}_3$ point along the intervening bond vectors, i.e., along the C_1^α—C_1', N_2—C_2^α, C_2'—C_2^α, and C_3^α—N_3 bonds; their directions are defined to point symmetrically inwards towards C_2^α (Fig. 3).

The atom C_2^α is restricted to a circle centered on the axis **r** (Fig. 2), because it must lie on a sphere of radius q_1 centered on C_1^α and on a sphere of radius q_3 centered on C_3^α. The intersection of these spheres is the circle on which C_2^α must be found.

Several angles between these unit vectors may be defined as well. The angles β_1, β_2, and β_3 represent the angles between the unit vectors $\hat{\mathbf{q}}_1$ and $\hat{\mathbf{r}}$, $\hat{\mathbf{q}}_1$, and $\hat{\mathbf{q}}_3$, and $\hat{\mathbf{q}}_3$ and $\hat{\mathbf{r}}$, respectively (Fig. 2). The angles γ_1 and γ_3 represent the angles between the unit vectors $\hat{\mathbf{q}}_1$ and $\hat{\mathbf{s}}_1$ and between $\hat{\mathbf{q}}_3$ and $\hat{\mathbf{p}}_3$, respectively, whereas the angle γ_2 is subtended by the unit vectors $\hat{\mathbf{q}}_3$ and $\hat{\mathbf{s}}_2$ (Fig. 3).

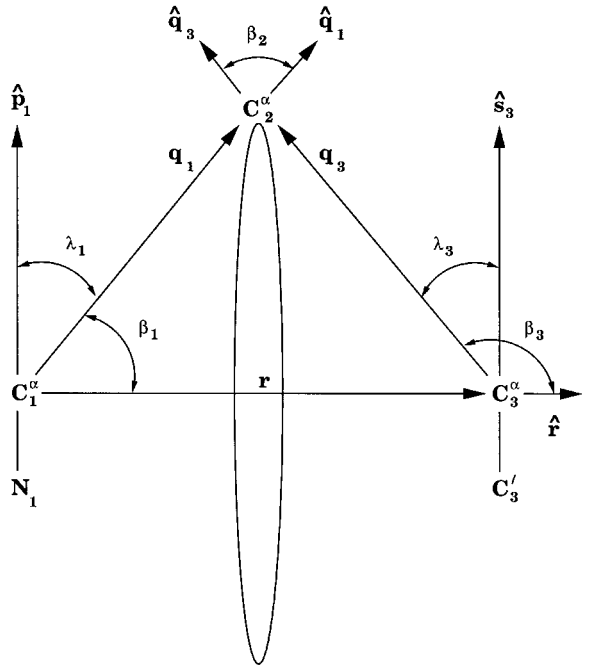

FIGURE 2. Illustration of the geometric definitions of the vectors $\mathbf{q}_1$ and $\mathbf{q}_3$ and of the angles β_1, β_2, β_3, λ_1, and λ_3. Because the vectors $\mathbf{q}_1$ and $\mathbf{q}_3$ have fixed length, the position of C_2^α is constrained to lie on a circle that is the intersection of two spheres of radii q_1 and q_3 whose centers are separated by the vector $\mathbf{r}$.

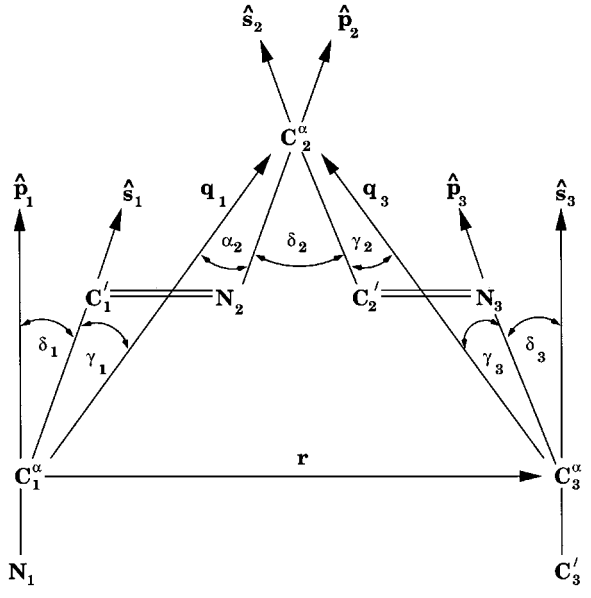

FIGURE 3. Illustration of the geometric definitions of the four unit vectors $\hat{\mathbf{s}}_1$, $\hat{\mathbf{p}}_2$, $\hat{\mathbf{s}}_2$, and $\hat{\mathbf{p}}_3$ and of the angles γ_1, δ_1, α_2, δ_2, γ_2, γ_3, and δ_3. All seven angles are fixed; δ_1, δ_2, and δ_3 are fixed by the bond angles at the C^α atoms, and the remaining four angles are fixed by the geometry of the peptide group.

Similarly, the angle α_2 is defined as the angle between the unit vectors $\hat{\mathbf{q}}_1$ and $\hat{\mathbf{p}}_2$ (Fig. 3). The angles δ_1 and δ_3 are the supplements of the bond angles at C_1^α and C_3^α, respectively, whereas δ_2 is simply the bond angle at C_2^α (Fig. 3). Last, the angles λ_1 and λ_3 are the angles between the unit vectors $\hat{\mathbf{q}}_1$ and $\hat{\mathbf{p}}_1$ and between $\hat{\mathbf{q}}_3$ and $\hat{\mathbf{s}}_3$, respectively (Fig. 2).

The key insight is that most of these angles are fixed and known in advance. The angles α_1 and α_3 are fixed by the dot products of the (known) unit vectors $\hat{\mathbf{p}}_1$ and $\hat{\mathbf{r}}$ and of $\hat{\mathbf{s}}_3$ and $\hat{\mathbf{r}}$, respectively (Fig. 1), while the angles β_1, β_2, and β_3 are fixed by the known magnitudes and coplanarity of the vectors $\mathbf{q}_1$, $\mathbf{q}_3$, and $\mathbf{r}$ (Fig. 2). Seven angles (γ_1, δ_1, α_2, γ_2, δ_2, γ_3, and δ_3) are fixed by the rigid bond geometry (Fig. 3). For example, the rigid ECEPP geometry[12] dictates that $\gamma_1 = \gamma_2 = 20.9°$ and $\alpha_2 = \gamma_3 = 14.9°$ (for *trans* peptide bonds) whereas δ_1, δ_2, and δ_3 are determined by the bond angles at the C^α atoms, whose values depend on the particular side chain (in ECEPP). Only λ_1 and λ_3 are variable, as may be seen by rotating C_2^α about its circle (Fig. 2). The constancy of these 12 angles suggests transferring the problem to the unit sphere, where it will be solved below.

By contrast, only the three unit vectors ($\hat{\mathbf{r}}$, $\hat{\mathbf{p}}_1$, and $\hat{\mathbf{s}}_3$) are known explicitly in advance. However, some general relationships among the vectors are known, especially concerning coplanarity. The unit vectors $\hat{\mathbf{r}}$, $\hat{\mathbf{q}}_1$, and $\hat{\mathbf{q}}_3$ lie in a common plane, because the corresponding vectors are linear combinations of one another, for example, $\mathbf{q}_1 = \mathbf{r} + \mathbf{q}_3$. Moreover, the planarity of the peptide group requires the coplanarity of the unit vectors $\hat{\mathbf{s}}_1$, $\hat{\mathbf{q}}_1$, and $\hat{\mathbf{p}}_2$, as well as that of the unit vectors $\hat{\mathbf{s}}_2$, $\hat{\mathbf{q}}_3$, and $\hat{\mathbf{p}}_3$. This coplanarity will prove useful in deriving the third biquadratic equation below.

Three Spherical Quadrangles and Their Solution

The unit vectors described in the previous section form three spherical quadrangles when transferred to the unit sphere, for example, the quadrangle formed by the unit vectors $\hat{\mathbf{r}}$, $\hat{\mathbf{q}}_1$, $\hat{\mathbf{s}}_1$, and $\hat{\mathbf{p}}_1$ (Figs. 4–6). The sides of the quadrangle have arc lengths equal to the angle (in radians) between the two corresponding unit vectors, and thus equal the angles α_1, β_1, γ_1, and δ_1, which are known and fixed (Figs. 1–3 and 6). Nevertheless, this quadrangle is not rigid, because the diagonal length λ_1 can vary (Fig. 2), and with it, the two interior spherical

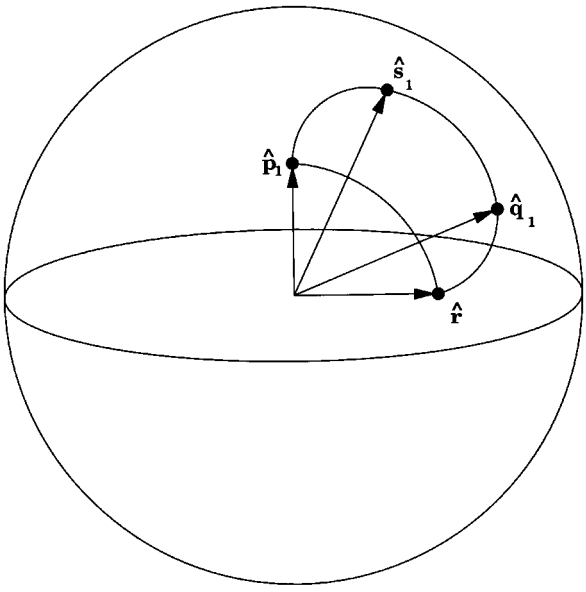

FIGURE 4. Illustration of the quadrangle formed on the unit sphere by the four unit vectors $\hat{\mathbf{r}}$, $\hat{\mathbf{q}}_1$, $\hat{\mathbf{p}}_1$, and $\hat{\mathbf{s}}_1$. Two other quadrangles will be formed by the unit vectors $\hat{\mathbf{r}}$, $\hat{\mathbf{q}}_3$, $\hat{\mathbf{p}}_3$, and $\hat{\mathbf{s}}_3$ and by the unit vectors $\hat{\mathbf{q}}_1$, $\hat{\mathbf{q}}_3$, $\hat{\mathbf{p}}_2$, and $\hat{\mathbf{s}}_2$.

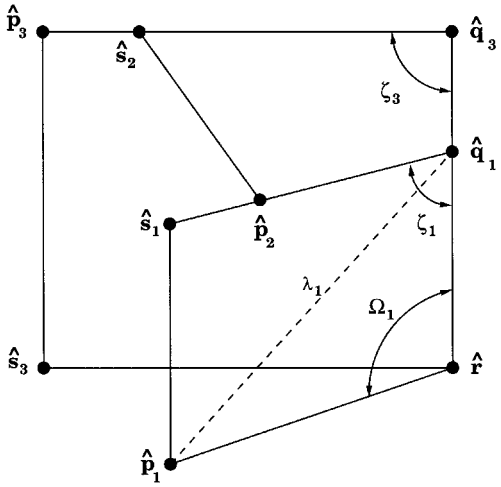

FIGURE 5. A projection of all the relevant unit vectors onto the unit sphere, where geodesics (great circles) are depicted by straight solid lines. The three quadrangles are evident in $\hat{\mathbf{r}}\hat{\mathbf{q}}_1\hat{\mathbf{s}}_1\hat{\mathbf{p}}_1$, $\hat{\mathbf{r}}\hat{\mathbf{q}}_3\hat{\mathbf{p}}_3\hat{\mathbf{s}}_3$, and in $\hat{\mathbf{q}}_1\hat{\mathbf{q}}_3\hat{\mathbf{s}}_2\hat{\mathbf{p}}_2$. The assumed planarity of the peptide group ensures that the unit vectors $\hat{\mathbf{s}}_1$, $\hat{\mathbf{p}}_2$, and $\hat{\mathbf{q}}_1$, as well as the unit vectors $\hat{\mathbf{p}}_3$, $\hat{\mathbf{s}}_2$, and $\hat{\mathbf{q}}_3$, are coplanar. The unit vectors $\hat{\mathbf{r}}$, $\hat{\mathbf{q}}_1$, and $\hat{\mathbf{q}}_3$ are coplanar by construction. The variable angle λ_1 is indicated by a dashed line.

angles Ω_1 and ζ_1 (Fig. 6). Appendix A derives in detail an equation relating these two spherical angles, which can be cast as a biquadratic equation

$$D_0 + D_1 w_1^2 + D_2 z_1^2 + D_3 w_1^2 z_1^2 + D_4 w_1 z_1 = 0 \quad (1)$$

where z_1 and w_1 are defined by the half-angle formulae

$$w_1 \equiv \tan\left(\frac{\Omega_1}{2}\right) \quad (2)$$

$$z_1 \equiv \tan\left(\frac{\zeta_1}{2}\right) \quad (3)$$

and the five coefficients D_0–D_4 depend only on the fixed angles α_1, β_1, γ_1, and δ_1.

Similarly, the unit vectors $\hat{\mathbf{r}}$, $\hat{\mathbf{q}}_3$, $\hat{\mathbf{p}}_3$, and $\hat{\mathbf{s}}_3$ form a spherical quadrangle when transferred to the unit sphere (Fig. 7). The sides of this quadrangle are fixed at α_3, β_3, γ_3, and δ_3, respectively, although this quadrangle is not rigid, because the diagonal length λ_3 can vary, and with it, the two interior spherical angles Ω_3 and ζ_3. The equation relating these two spherical angles can be taken directly from the solution of Appendix A. Appendix B shows how the spherical angle Ω_3 can be eliminated in favor of Ω_1, yielding the second biquadratic equation

$$P_{00} + P_{10} w_1 + P_{01} z_3 + P_{11} w_1 z_3 + P_{20} w_1^2 + P_{02} z_3^2 + P_{12} w_1 z_3^2 + P_{21} w_1^2 z_3 + P_{22} w_1^2 z_3^2 = 0 \quad (4)$$

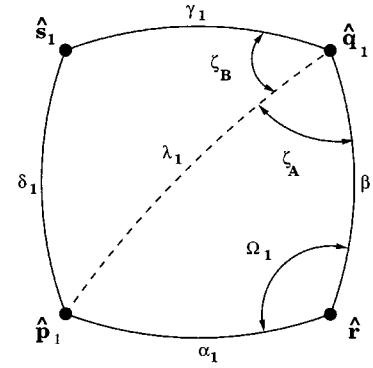

FIGURE 6. The first quadrangle on the unit sphere, the analogue of Figure 4. The first of three biquadratic equations is derived from this quadrangle, relating the spherical angle $\zeta_1 \equiv \zeta_A + \zeta_B$ (see Fig. 5) to the spherical angle Ω_1 and the fixed quantities α_1, β_1, γ_1, and δ_1 (depicted with solid lines). The variable angle λ_1 is depicted by a dashed line.

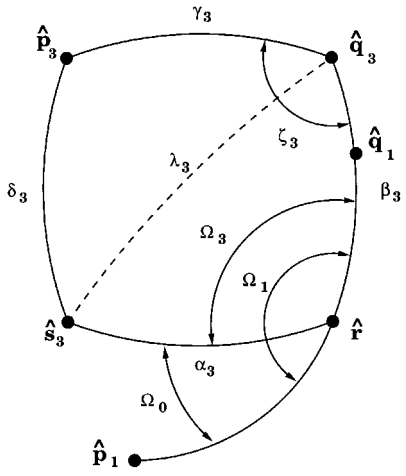

FIGURE 7. The second quadrangle on the unit sphere, formed by the unit vectors $\hat{\mathbf{r}}$, $\hat{\mathbf{q}}_3$, $\hat{\mathbf{p}}_3$, and $\hat{\mathbf{s}}_3$. The second biquadratic equation is derived from this quadrangle, relating the spherical angle ζ_3 to the spherical angle Ω_1 and the fixed quantities α_3, β_3, γ_3, δ_3 (depicted with solid lines) and the fixed spherical angle Ω_0. The variable angle λ_3 is depicted by a dashed line.

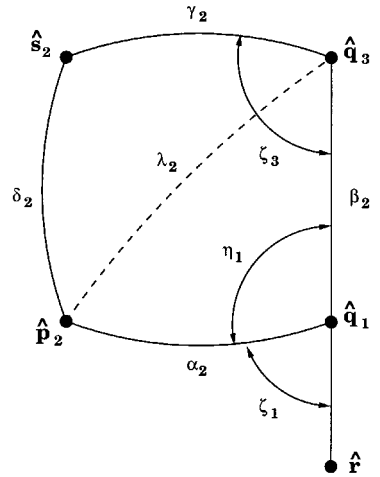

FIGURE 8. The third quadrangle on the unit sphere, formed by the unit vectors $\hat{\mathbf{q}}_1$, $\hat{\mathbf{q}}_3$, $\hat{\mathbf{s}}_2$, and $\hat{\mathbf{p}}_2$. The third biquadratic equation is derived from this quadrangle, relating the spherical angle ζ_1 to the spherical angle ζ_3 and the fixed quantities α_2, β_2, γ_2, and δ_2. The variable angle λ_2 is shown by a dashed line, by analogy with its two predecessors (Figs. 6 and 7).

where z_3 is likewise defined by the half-angle formula

$$z_3 \equiv \tan\left(\frac{\zeta_3}{2}\right) \tag{5}$$

and the nine coefficients P_{00}–P_{22} depend only on the fixed angles α_3, β_3, γ_3, δ_3, and Ω_0, which is the fixed spherical angle $\angle\hat{\mathbf{p}}_1\hat{\mathbf{r}}\hat{\mathbf{s}}_3$.

The unit vectors $\hat{\mathbf{q}}_1$, $\hat{\mathbf{q}}_3$, $\hat{\mathbf{s}}_2$, and $\hat{\mathbf{p}}_2$ form a third spherical quadrangle when transferred to the unit sphere (Fig. 8). The sides of this quadrangle are fixed at α_2, β_2, γ_2, and δ_2, respectively. Like its predecessors, this quadrangle is not rigid, and the two interior angles η_1 and ζ_3 can vary. The equation relating these two spherical angles can be taken directly from the solution of Appendix A. Appendix C shows how the spherical angle η_1 can be eliminated in favor of ζ_1, yielding the third biquadratic equation

$$M_0 + M_1 z_1^2 + M_2 z_3^2 + M_3 z_1^2 z_3^2 + M_4 z_1 z_3 = 0 \tag{6}$$

where the five coefficients M_0–M_4 depend only on the fixed angles α_2, β_2, γ_2, and δ_2.

The three simultaneous biquadratic eqs. (1), (4), and (6) in three variables (w_1, z_1, and z_3) can be solved using the method of resultants (Appendix D), resulting in a polynomial equation of degree 16 in w_1

$$\begin{aligned} R_0 &+ R_1 w_1 + R_2 w_1^2 + R_3 w_1^3 + R_4 w_1^4 \\ &+ R_5 w_1^5 + R_6 w_1^6 R_7 w_1^7 + R_8 w_1^8 + R_9 w_1^9 \\ &+ R_{10} w_1^{10} + R_{11} w_1^{11} + R_{12} w_1^{12} + R_{13} w_1^{13} \\ &+ R_{14} w_1^{14} + R_{15} w_1^{15} + R_{16} w_1^{16} = 0 \end{aligned} \tag{7}$$

where the 17 coefficients R_0–R_{16} are determined by the fixed angles of the problem. Because such a polynomial can have at most 16 real roots, and because each real root corresponds to one valid loop closure (as shown in Appendix D), there are at most 16 solutions for the closing of a tripeptide loop. Practically, the number may be smaller; the empirically observed limit is that a tripeptide loop-closure problem can have at most eight solutions.[1] Appendix E details how to find the atomic positions for each real root of the loop closure polynomial.

Identifying Feasible Loop Closure Constraints

The previous section showed that tripeptide loop closures correspond to the real roots of a polynomial of degree 16. This section computes bounds on the valid solutions and checks the feasibility of the given loop-closure constraints, i.e., whether the loop-closure problem has any solu-

tions given its particular values for the fixed angles. Similar criteria were formulated in the original article to show that cyclic tri- and tetrapeptides could not be formed with *trans* peptide bonds.[1]

As shown in Figure 3, the two virtual bonds $\mathbf{q}_1$ and $\mathbf{q}_3$ are fixed in length, and the angle β_2 between them is bounded between $\delta_2 + \alpha_2 + \gamma_2$, and $\delta_2 - \alpha_2 - \gamma_2$, respectively. Thus, the distance r between C_1^α and C_3^α is likewise bounded between two limits, given by the law of cosines

$$r_{\max}^2 = q_1^2 + q_3^2 - 2q_1q_3\cos(\delta_2 + \alpha_2 + \gamma_2) \quad (8)$$

$$r_{\min}^2 = q_1^2 + q_3^2 - 2q_1q_3\cos(\delta_2 - \alpha_2 - \gamma_2) \quad (9)$$

For a given loop-closure problem, if r lies outside this allowed range, there can be no solutions. This is a geometrical derivation of the bounds found numerically in Eq. (52) of the original article.[1]

The quadratic form of the three biquadratic eqs. (1), (4), and (6) can be exploited to obtain bounds on their solutions, i.e., the spherical angles Ω_1, ζ_1, and ζ_3. For illustration, consider the first biquadratic eq. (1). Gathering terms yields a simple quadratic equation in z_1

$$(D_0 + D_1w_1^2) + (D_4w_1)z_1 + (D_2 + D_3w_1^2)z_1^2 = 0 \quad (10)$$

whose solution is given by the quadratic formula

$$z_1 = \frac{1}{2(D_0 + D_1w_1^2)} \times \left[-D_4w_1 \pm \sqrt{(D_4w_1)^2 - 4(D_0 + D_1w_1^2)(D_2 + D_3w_1^2)} \right] \quad (11)$$

Because the solution for z_1 must be real, the discriminant under the radical must be positive, yielding a quadratic inequality in w_1^2

$$-4D_1D_3w_1^4 + (D_4^2 - 4D_0D_3 - 4D_1D_2)w_1^2 - 4D_0D_2 \geq 0 \quad (12)$$

Depending on the values of the coefficients D_0–D_4, this inequality restricts Ω_1 to zero, one, two or three intervals, which may be easily determined from the roots of the discriminant, which can be found by again applying the quadratic formula. Likewise, eq. (4) can also be cast as a quadratic equation in z_3, and another set of bounds on w_1 determined. If either equation yields no valid solutions for w_1, or if the two sets of intervals do not overlap, then the loop-closure problem has no solution. Similar pairs of bounds can be imposed on the spherical angles ζ_1 and ζ_3. Such bounds bracket the solutions and may be useful starting points for a numerical solution of the three biquadratic equations.

In addition to these preliminary checks, the method of Sturm chains[35,36] may be used to determine the number of real roots of the polynomial (7) and, thus, the exact number of valid loop closures. Sturm chains give the number of real roots of a polynomial in any interval, including the whole real line $(-\infty, \infty)$, without requiring that those roots be determined. Thus, the number of exact loop closures may be computed without finding any roots, once the coefficients of the loop-closure polynomial (7) have been computed. The method of Sturm chains is illustrated in Appendix F.

The General Seven-Atom Loop-Closure Problem

The solution of the tripeptide loop-closure problem does not depend on any peculiarity of polypeptide chains, and therefore, may be applied to the general problem of closing a loop of nine atoms with fixed bond lengths and bond angles and three distance constraints corresponding to the distance r between C_1^α and C_3^α, the distance q_1 between C_1^α and C_2^α, and the distance q_3 between C_2^α and C_3^α.

The same solution also applies to the general problem of closing a loop of seven atoms with only one distance constraint (Fig. 9). The loop-closure constraints can be depicted as angular constraints between unit vectors and, moreover, these unit vectors form three coupled quadrangles when transferred to the unit sphere, as in the tripeptide loop-closure problem. As depicted in Figure 9, let the atoms of the loop be numbered from 1 to 7, where atoms 1, 2, 6, and 7 are fixed in space and the positions of atoms 3–5 remain to be found. Let the bond vectors be denoted by $\mathbf{g}_1$–$\mathbf{g}_3$ and $\mathbf{h}_1$–$\mathbf{h}_3$ (Fig. 9), where the angles between adjacent bond vectors are fixed quantities. Similarly, let the vectors joining atoms 2 and 6 to atom 4 be denoted by $\mathbf{q}_1$ and $\mathbf{q}_3$, respectively; the lengths of these two vectors can be found from the fixed lengths of their component bond vectors and the fixed bond

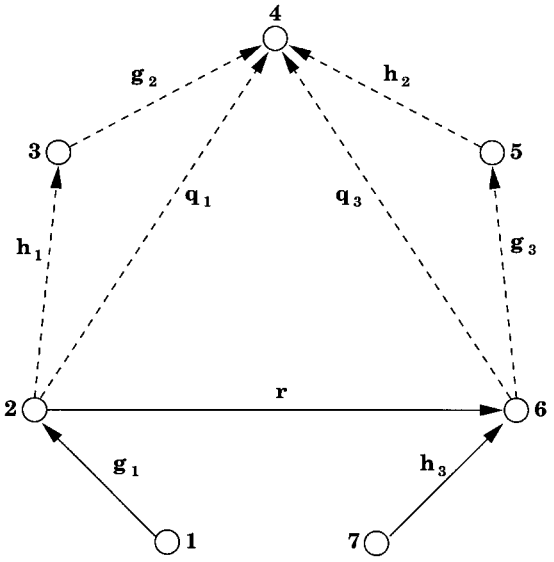

FIGURE 9. Illustration of the general seven-atom loop-closure problem. The atoms are numbered from 1 to 7, of which 1, 2, 6, and 7 are fixed, while the positions of atoms 3–5 must be determined. The successive bond vectors are labeled $\mathbf{g}_1$, $\mathbf{h}_1$, $\mathbf{g}_2$, $\mathbf{h}_2$, $\mathbf{g}_3$, and $\mathbf{h}_3$, by analogy with the corresponding vectors **p** and **s** in Figure 1. The vectors $\mathbf{q}_1$, $\mathbf{q}_3$, and **r** are likewise labeled to agree with the corresponding vectors in Figures 1 and 2.

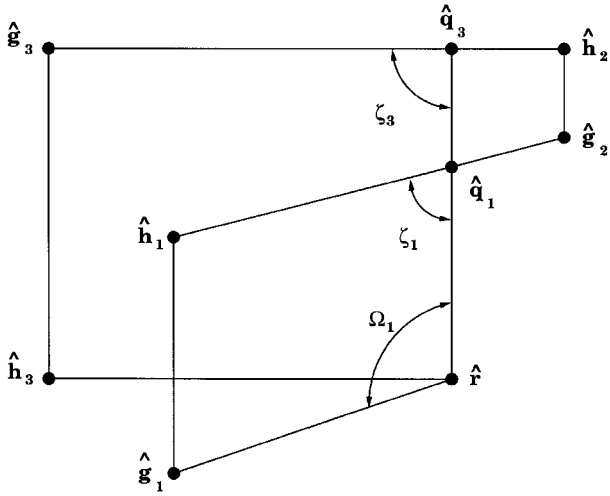

FIGURE 10. The unit vectors of Figure 9 form three quadrangles when transferred to the unit sphere, by analogy with Figure 5. As in that figure, the points represent their respective unit vectors and the solid lines represent arcs of fixed angle. These three quadrangles are equivalent to three coupled biquadratic equations, which may be solved as in the tripeptide loop-closure problem.

angle between them. The angles between $\mathbf{q}_1$ and $\mathbf{q}_3$ and their component bond vectors can also be determined. Last, the angles between **r**, $\mathbf{q}_1$, and $\mathbf{q}_3$ can be computed from their known lengths. When transferred to the unit sphere, these angular constraints make three spherical quadrangles (Fig. 10), which are equivalent to those of the tripeptide loop closure problem.

An Eight-Atom Loop-Closure Problem

As shown in the previous section, the fixed peptide bond dihedral angles and the resulting distance constraints between the atoms C_1^α and C_2^α and between C_3^α and C_2^α render the (nine-atom) tripeptide loop-closure problem equivalent to the general seven-atom loop-closure problem. An interesting intermediate problem is that of the eight-atom ring with two constraints, illustrated here by the disulfide bond in proteins.

The disulfide loop-closure problem is presented in Figure 11.[37,38] The atoms in the two halves of the disulfide loop are labeled A and B. The four atoms N_A, C_A^α, C_B^α, and C'_B are held fixed, while

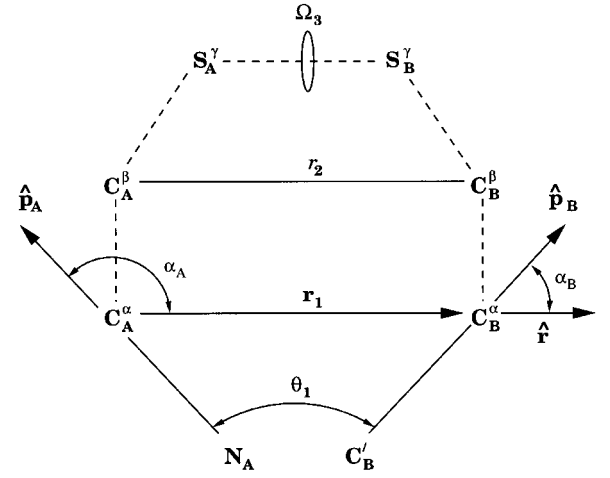

FIGURE 11. Illustration of the (eight-atom) disulfide loop-closure problem. The positions of four atoms (N_A, C_A^α, C_B^α, and C'_B) are fixed, while those of four other atoms (C_A^β, S_A^γ, S_B^γ, and C_B^β) remain to be determined. The eight-atom loop requires one additional constraint, compared to the seven-atom loop, which is taken here as the value of the dihedral angle Ω_3, which sets the value of the distance r_2. The angles α_A and α_B are fixed by the positions of the four fixed atoms.

the positions of four other atoms ($C_A^\beta, S_A^\gamma, S_B^\gamma, C_B^\beta$) remain to be determined. The two constraints for this problem are the vector $\mathbf{r}_1$ between the two fixed C^α atoms and the dihedral angle Ω_3, which fixes the distance r_2 between the two C^β atoms, whose positions are initially undetermined. It should be noted that the vector $\mathbf{r}_2$ is not given, only its length r_2; this second constraint is chosen because the dihedral angle Ω_3 is generally confined near $\pm 90°$. The cysteine residues are not required to be adjacent in the amino acid sequence. However, the backbone conformation at C_A^α and C_B^α cannot be completely defined; otherwise, the positions of the two C^β atoms would be fixed, causing the problem to be overdetermined and (in general) to have no solution.

The geometric definitions are given in Figures 11–14. Once again, several angles between unit vectors are fixed in advance, although the unit vectors themselves may be unknown. The known unit vectors $\hat{\mathbf{p}}_A$ and $\hat{\mathbf{p}}_B$ represent the unit bond vectors between the fixed atoms N_A and C_A^α and between C'_B and C_B^α; the angle between these two unit vectors is denoted θ_1, while their respective angles with the (known) unit vector $\hat{\mathbf{r}}$ are α_A and α_B (Fig. 11). The unknown bond unit vectors are represented by $\hat{\mathbf{q}}_A$, $\hat{\mathbf{q}}_B$, $\hat{\mathbf{s}}_A$, and $\hat{\mathbf{s}}_B$. The fixed bond angles between $\hat{\mathbf{p}}_A$ and $\hat{\mathbf{q}}_A$ and between $\hat{\mathbf{p}}_B$ and $\hat{\mathbf{q}}_B$ are denoted β_A and β_B, respectively, while γ_A and γ_B represent the angles between $\hat{\mathbf{q}}_A$ and $\hat{\mathbf{s}}_A$ and between $\hat{\mathbf{q}}_B$ and $\hat{\mathbf{s}}_B$, respectively. The angles δ_A and δ_B represent the respective angles between the unit vectors $\hat{\mathbf{s}}_A$ and $\hat{\mathbf{s}}_B$ with the vector $\hat{\mathbf{r}}_2$, while ϵ_A and ϵ_B represent their respective angles with the vector $\hat{\mathbf{r}}_3$. Finally, the angle between the two unit vectors $\hat{\mathbf{s}}_A$ and $\hat{\mathbf{s}}_B$ is denoted by θ_2.

The (initially unknown) angles δ_A, δ_B, and θ_2 can be determined as follows. The quadrangle in Figure 13 can be expressed as a vector equation

$$\mathbf{r}_3 - \mathbf{s}_B = \mathbf{r}_2 - \mathbf{s}_A \tag{13}$$

which, when squared, gives the relation

$$r_3^2 + s_B^2 - 2r_3 s_B \cos \epsilon_B = r_2^2 + s_A^2 - 2r_2 s_A \cos \delta_A \tag{14}$$

Thus, the angle δ_A can be determined from other fixed quantities

$$\cos \delta_A = \left[r_2^2 + s_A^2 - r_3^2 - s_B^2 + 2r_3 s_B \cos \epsilon_B \right] \times \left(\frac{1}{2r_2 s_A} \right) \tag{15}$$

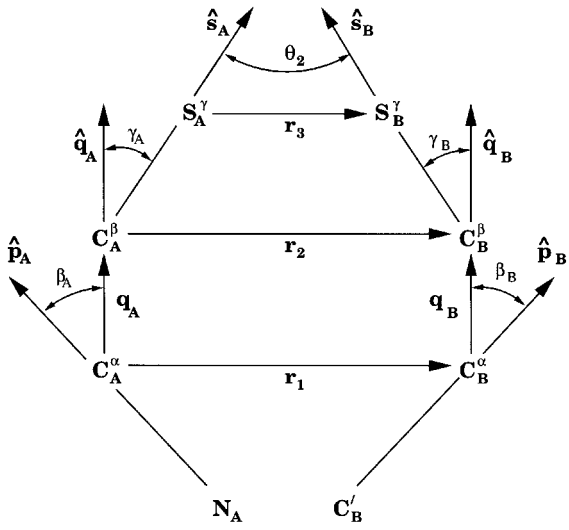

FIGURE 12. The definitions of several vectors and the angles between them. The bond unit vectors being sought are denoted by $\hat{\mathbf{q}}_A$, $\hat{\mathbf{q}}_B$, $\hat{\mathbf{s}}_A$, and $\hat{\mathbf{s}}_B$. The (fixed) angles between these vectors are β_A, β_B, γ_A, γ_B, and θ_2. Additional angles are defined in Figure 13.

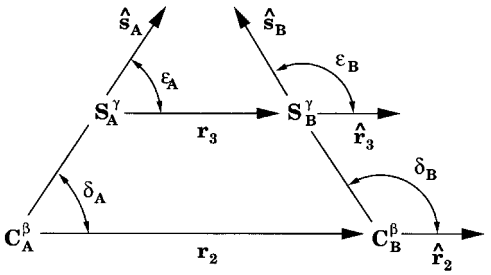

FIGURE 13. The C^β—S^γ bond unit vectors $\hat{\mathbf{s}}_A$ and $\hat{\mathbf{s}}_B$ make four (fixed) angles with the vectors $\hat{\mathbf{r}}_2$ and $\hat{\mathbf{r}}_3$, namely δ_A, δ_B, ϵ_A, and ϵ_B, respectively.

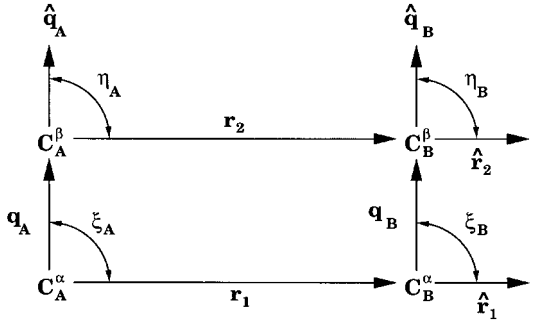

FIGURE 14. The central quadrangle of the disulfide loop-closure problem, composed of the four vectors $\mathbf{q}_A$, $\mathbf{q}_B$, $\mathbf{r}_1$, and $\mathbf{r}_2$, where four variable angles ξ_A, ξ_B, η_A, and η_B are defined for convenience.

A similar equation can be obtained for δ_B by writing the vector equation [eq. (13)] in the form

$$\mathbf{r}_3 + \mathbf{s}_A = \mathbf{r}_2 + \mathbf{s}_B \tag{16}$$

which, when squared, gives the relation

$$r_3^2 + s_A^2 + 2r_3 s_A \cos \epsilon_A = r_2^2 + s_B^2 + 2r_2 s_B \cos \delta_B \tag{17}$$

Last, the angle θ_2 between the two vectors $\mathbf{s}_A$ and $\mathbf{s}_B$ can be determined by writing the vector equation [eq. (13)] in the form

$$\mathbf{r}_2 = \mathbf{r}_3 + \mathbf{s}_A - \mathbf{s}_B \tag{18}$$

which, when squared, gives the relation for θ_2

$$r_2^2 = r_3^2 + s_A^2 + s_B^2 + 2r_3 s_A \cos \epsilon_A - 2r_3 s_B \cos \epsilon_B - 2s_A s_B \cos \theta_2 \tag{19}$$

Thus, all 12 angles (α_A, α_B, β_A, β_B, γ_A, γ_B, δ_A, δ_B, ϵ_A, ϵ_B, θ_1, and θ_2) are fixed at known values. The angles α_A, α_B, and θ_1 are determined by the four fixed atoms C_A^α, N_A, C_B^α, and C_B' (Fig. 11). The angles δ_A, δ_B, and θ_2 are fixed by the equations above, and the remaining angles are fixed by the bond angles (Figs. 12 and 13).

Transferring the unit vectors onto the unit sphere gives the diagram of Figure 15. The dashed lines indicate the four angles (ξ_A, ξ_B, η_A, and η_B) of the central quadrangle (Fig. 14), which are not fixed by the geometry, but are determined by the distance constraints. The eight-atom loop-closure problem differs from the seven-atom and nine-atom loop-closure problems in that *not every constraint of the eight-atom loop-closure problem can be expressed as a fixed angular constraint*. In other words, the defining equations involve lengths as well as angles, and thus, the problem cannot be solved completely on the unit sphere. Figure 15 also introduces four spherical angles (Ω_A, Ω_B, ζ_A, and ζ_B) that will prove essential.

The solution is obtained as follows (Appendix G). A biquadratic equation relating the spherical angles Ω_A to Ω_B is derived first, followed by the analogous equation relating the spherical angles ζ_A and ζ_B. These equations depend on the distances r_1 and r_2. Once these distances are fixed, the conformation of the central quadrangle consisting of the vectors $\hat{\mathbf{r}}_1$, $\hat{\mathbf{r}}_2$, $\hat{\mathbf{q}}_A$, and $\hat{\mathbf{q}}_B$ has two degrees of freedom, which can be taken as the angles η_A and η_B (Fig. 14). Applying the spherical law of cosines to these two angles yields two additional biquadratic equations relating the spherical angles

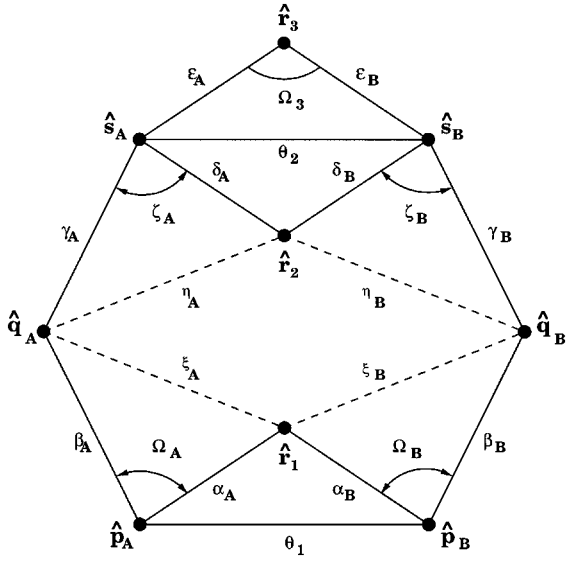

FIGURE 15. The unit vectors of the disulfide loop-closure problem projected onto the unit sphere. As in Figures 5 and 10, the solid lines represent arcs of fixed angles. The dashed lines indicate arcs of variable angle whose values will be determined by the constraints of the problem. This loop-closure problem will be solved by finding four equations for the four spherical angles Ω_A, Ω_B, ζ_A, and ζ_B defined in this figure.

ζ_B to Ω_A and ζ_A to Ω_B, respectively. This system of four biquadratic equations in four variables (corresponding to Ω_A, Ω_B, ζ_A, and ζ_B) can again be solved by the method of resultants, which yields a polynomial equation in one variable of degree 32. The real roots of this polynomial allow one to solve for all valid loop closures. Appendix G gives the details of the derivation.

Closing Longer Loops

The above sections describe how to close short loops such as the tripeptide and disulfide-bonded loops. However, protein simulations often seek to close longer loops, typically ranging from 5–15 residues. One application is the homology modeling of proteins, where the hydrophobic core is generally conserved but significant structural changes occur in the surface loops. This section discusses methods for closing such longer loops.

A key difficulty is that longer loop closure problems are usually underdetermined, and thus additional constraints must be imposed to obtain specific solutions. A loop closure problem of n atoms generally requires $n - 6$ constraints (in addition to

the constraints of rigid bond lengths and bond angles) to determine specific solutions.[1] Thus, a polypeptide loop closure problem involving r amino acid residues (where the initial and final C^{α} atoms are fixed) generally requires $2r - 6$ additional constraints before specific loop conformations can be generated; each amino acid has two undetermined dihedral angles ϕ and ψ, while six of these dihedral angles are determined by the Gō–Scheraga equations. For illustration, a pentapeptide ($r = 5$) loop-closure problem has $2r - 6 = 4$ extra degrees of freedom[7] while a 15-residue ($r = 15$) loop-closure problem has $2r - 6 = 24$ extra degrees of freedom. An equal number of constraints must be imposed to produce specific loop closures.

There are several strategies for imposing such constraints. Let m represent the number of constraints that must be added to find specific loop closures. A brute-force approach is to vary m dihedral angles systematically and exhaustively (grid search), as recommended in the original article.[1] In principle, any m dihedral angles may be chosen,[1,18] but it is convenient to vary all but the middle six dihedral angles, which can then be found using the tripeptide loop-closure method. For example, the closure of a pentapeptide loop requires four extra constraints, which may be chosen as the initial two and final two dihedral angles; these angles can be varied exhaustively and, for each selected set of the four values, the inner six dihedral angles can be determined by solving the middle tripeptide loop-closure problem,[7] for example, by the method described above (Fig. 16). The distance and angle feasibility criteria given above and the Sturm chain method can check quickly whether the given values of the outer dihedral angles permit any loop closures.

Such exhaustive brute-force sampling does not work well for very long loops, however, because the number of combinations grows exponentially with the number of free dihedral angles. For example, if the four free dihedral angles of the pentapeptide loop are varied in steps of 3 degrees, then $(360/3)^4 \approx 2 \times 10^8$ different tripeptide loop-closure problems must be solved. For 24 free dihedral angles, the number of exhaustive combinations is approximately 8×10^{49}, an astronomical figure. This problem can be alleviated somewhat by stochastic (rather than exhaustive) sampling of the free dihedral angles. For example, a nonapeptide loop can be closed by randomly varying the dihedral angles of the three initial and three final

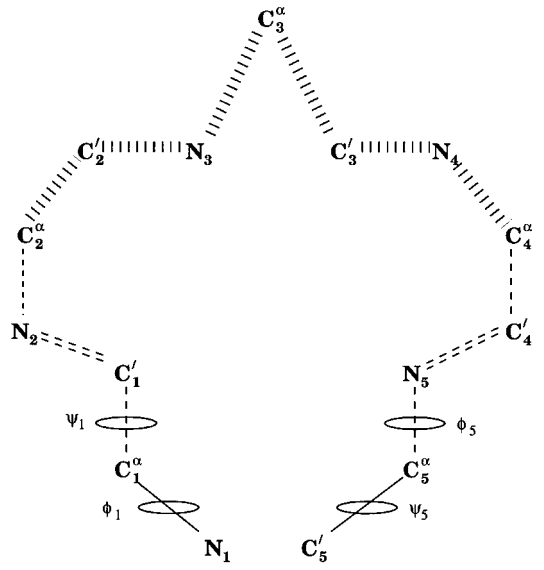

FIGURE 16. The brute-force method for closing longer loops is applied to a pentapeptide loops. The four dihedral angles ϕ_1, ψ_1, ϕ_5, and ψ_5 are sampled exhaustively. Holding these dihedral angles fixed determines the positions of six atoms C'_1, N_2, C^{α}_2, N_5, C'_4, and C^{α}_4. The remaining atoms can then be determined by solving the central tripeptide loop closure problem using the method described in the text.

residues, and solving the central tripeptide loop-closure problem exactly. However, a fundamental inefficiency of the brute-force approach is that most choices of the free dihedral angles do not bring the ends close enough together to allow the central tripeptide loop to close; the probability of randomly generating a valid conformation of the free dihedral angles decays rapidly as the loop length increases.[39,40]

A second approach to the closure of long loops is that of decimation.[40,41] In this approach, the long loop is represented by rigid fragments (remnants) of the loop, for example, every third peptide group. Given the positions of these remnants, the intervening atoms can be located by solving the independent loop-closure problems between them. As a simple example, consider the hexapeptide loop, where the first two and final two atoms (N_1, C^{α}_1, C^{α}_6, and C'_6) are fixed in space (Fig. 17). This hexapeptide loop-closure problem can be reduced to two independent tripeptide loop closure problems by decimation, i.e., by sampling the position of the fourth peptide group, which is equivalent to sampling the positions of the atoms C^{α}_3, C'_3, N_4, and C^{α}_4. Such sampling can be carried out in several ways, for example, by Monte Carlo or by

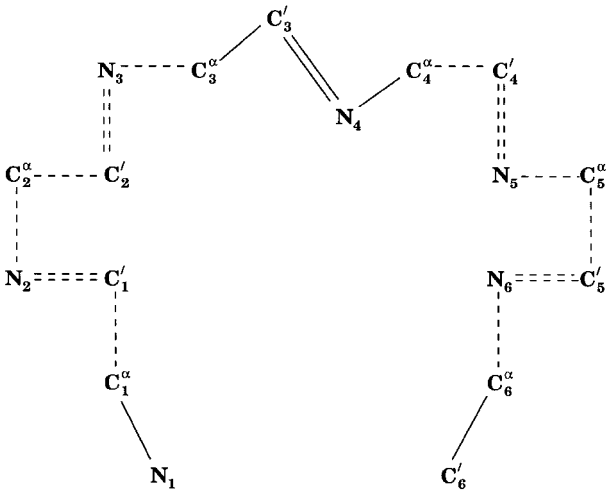

FIGURE 17. Illustration of the decimation method to close a hexapeptide loop. In this problem, the four atoms N_1, C^α_1, C^α_6, and C'_6 are fixed, and the position and orientation of the central peptide group is sampled, which is equivalent to sampling positions for the four atoms C^α_3, C'_3, N_4, and C^α_4. The positions of the remaining atoms can be found by applying the tripeptide loop closure algorithm (described in the text) twice, to the tripeptide C^α_1—C^α_2—C^α_3, and to the tripeptide C^α_4—C^α_5—C^α_6.

using experimentally derived constraints on the position of the remnant. Holding the remnant atoms in their sampled positions, the positions of the remaining backbone atoms can be found by solving the tripeptide loop-closure problems between C^α_1 and C^α_3 and between C^α_4 and C^α_6, respectively. By sampling different positions of the fourth peptide group, the full range of hexapeptide loop closures can be mapped out. Clearly, this method can be extended to longer loops, for example, to a nonapeptide by sampling the positions of the fourth and seventh peptide groups and solving the three loop-closure problems: between C^α_1 and C^α_3, between C^α_4 and C^α_6, and between C^α_7 and C^α_9, respectively. However, very long loops may have a small probability of being solved by the decimation method, if they are composed of many independent loop-closure problems, because all of these must be simultaneously soluble for the whole loop to close. The decimation approach seems best suited to situations in which a remnant is known to be positioned within a restricted region, for example, in a flexible docking problem where the positions of a few remnants of the ligand have been determined experimentally or from simulations.

A third approach to the closure of long loops is to apply the loop-closing algorithm hierarchically. Instead of sampling dihedral angles (as in the brute-force approach) or the positions and orientations of rigid fragments (as in the decimation approach), additional constraints can be imposed by sampling distances between atoms. Long loop-closure problems can then be solved while holding the distances between these atoms fixed at their sampled values. These constraints will be called ''pseudobonds,'' to distinguish them from covalent bonds and from virtual bonds, whose technical meaning is already defined.[34]

The heptapeptide loop-closure problem illustrates this approach (Fig. 18). The seven C^α atoms are connected by rigid virtual bonds of length 3.8 Å (assuming *trans* peptide groups). If the seven pair distances N_1—C^α_2, C^α_1—C^α_3, C^α_2—C^α_5, C^α_3—C^α_5,

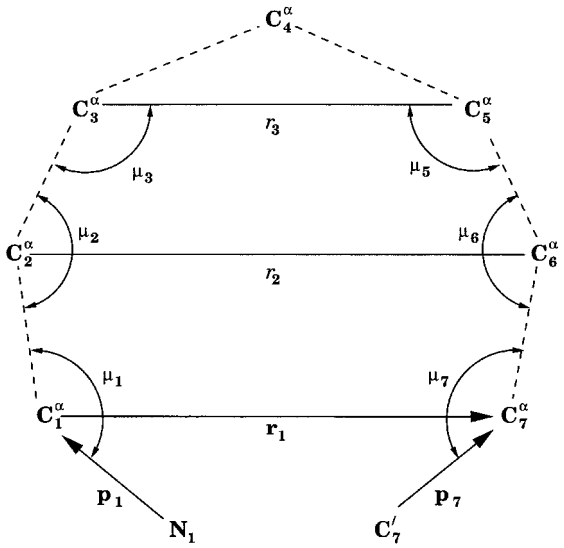

FIGURE 18. Even longer loops can be closed by applying the loop-closure algorithm hierarchically. A heptapeptide loop can be formed by closing the loop composed of N_1, C'_7, and the six C^α atoms of residues 1, 2, 3, 5, 6, and 7. The virtual bond lengths between adjacent C^α atoms (indicated by dashed lines) are fixed when the peptide bond dihedral angle is fixed, but the virtual bond angles are not, and must be sampled. With fixed virtual bond lengths and bond angles, the four unknown C^α positions may be determined using the eight-atom loop-closure method described in the text for disulfides. The positions of the atoms in the peptide groups between them can be determined by the method of Purisima and Scheraga.[42]. Finally, the central tripeptide loop-closure problem can be solved to find the positions of the atoms between C^α_3 and C^α_5. In all, 17 atoms have been positioned relative to the four fixed atoms (N_1, C^α_1, C^α_7, and C'_7).

C_3^α—C_6^α, C_5^α—C_7^α, and C_6^α—C_7' are sampled and held fixed, the bond angles μ_1—μ_7 are determined and the problem is reduced to solving an eight-atom loop-closure problem, in which the positions of the four atoms C_2^α, C_3^α, C_5^α, and C_6^α remain to be found. Once the positions of these four C^α atoms have been found, the positions of the atoms making up the 1–2, 2–3, 5–6, and 6–7 peptide groups can be determined, using the (covalent) bond angle constraints at C_1^α, C_2^α, C_6^α, and C_7^α, respectively.[42] Once all of these atoms have been positioned, the positions of the atoms between C_3^α and C_5^α can be determined by a normal tripeptide loop-closure method. Thus, we apply the eight-atom loop-closure algorithm to find the positions of all the C^α atoms except C_4^α, which is subsequently found using the normal tripeptide loop-closure algorithm. The seven-atom loop-closure algorithm can be applied analogously to solve the hexapeptide loop-closure algorithm.

A hierarchical approach can be used to solve even larger loops. Consider a 17-residue loop-closure problem where the initial two and final two atoms (N_1, C_1^α, C_{17}^α, and C_{17}') are fixed in space, and the positions of the intervening atoms are sought (Fig. 19). This problem can be solved by establishing pseudobonds between every fourth C^α atom, for example, between C_1^α and C_5^α, between C_4^α and C_9^α, and so on; the lengths of these pseudobonds and the angles between them can be sampled stochastically. Holding the sampled lengths and angles fixed, one can solve for the positions of the three atoms C_5^α, C_9^α, and C_{13}^α using the seven-atom loop-closure method derived above. The positions of the remaining atoms can be found in several ways, once the positions of these C^α atoms are fixed; some additional sampling is necessary, however, because the remaining positions are still underdetermined. One possibility is to sample the local dihedral angles at each of the five C^α atoms (C_1^α, C_5^α, C_9^α, C_{13}^α, and C_{17}^α), as well as the overall orientation at each C^α atom. This determines the positions of the atoms in eight peptide groups (C_1^α—C_2^α, C_4^α—C_5^α, C_5^α—C_6^α, C_8^α—C_9^α, C_9^α—C_{10}^α, C_{12}^α—C_{13}^α, C_{13}^α—C_{14}^α, and C_{16}^α—C_{17}^α) and, in effect, reduces the problem to a decimation loop closure. The tripeptide loop-closure algorithm may then be applied four times to find the positions of the remaining atoms, namely the atoms between C_2^α and C_4^α, between C_6^α and C_8^α, between C_{10}^α and C_{12}^α, and between C_{14}^α and C_{16}^α. Thus, the loop-closure algorithms have been applied hierarchically, first at a coarse scale to locate valid positions for the three C^α atoms, and later at a fine scale, to position the remaining atoms.

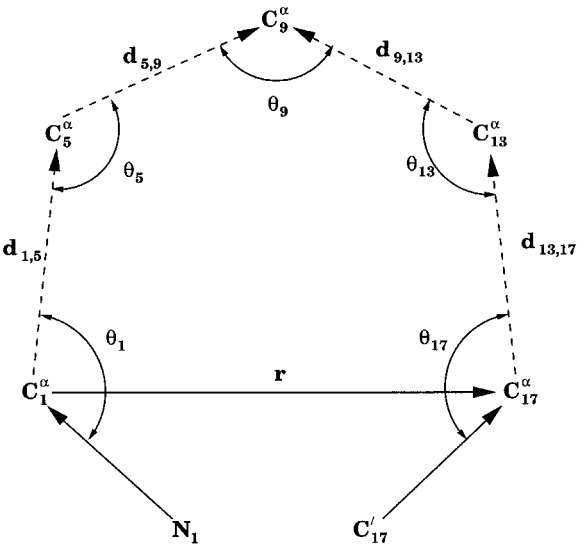

FIGURE 19. Closure of a 17-residue loop by applying the loop-closure algorithms hierarchically. In this problem, the pseudobond distances as well as the pseudobond angles must be sampled, because they are not fixed by the covalent bond geometry. Holding them fixed at the sampled values, the three C^α atoms of residues 5, 9, and 13 can be found from the positions of the four fixed atoms (N_1, C_1^α, C_{17}^α, and C_{17}'), using the general seven-atom loop-closure algorithm described in the text. The remaining atoms can be found by various methods, for example, by the tripeptide loop-closure method, as described in the text.

As noted above, all methods require some sampling to provide the m additional constraints necessary to close longer (underdetermined) loops exactly. In the brute-force approach, the free dihedral angles are sampled, for example, the initial and final two dihedral angles of the pentapeptide loop closure problem (Fig. 16). In the hierarchical approach, the distances between the C^α atoms are sampled, for example, the nine pair distances in Figure 19, N_1—C_5^α, C_1^α—C_5^α, C_1^α—C_9^α, C_5^α—C_9^α, C_5^α—C_{13}^α, C_9^α—C_{13}^α, C_9^α—C_{17}^α, C_{13}^α—C_{17}^α, and C_{13}^α—C_{17}'. In the decimation method, the positions of the remnants must be sampled, for example, the position of the central peptide group in Figure 17. It should be stressed that these variables need not be sampled completely at random, but that prior information may be used, for example, statistical distributions of the dihedral angles in the brute-force method, the statistical distribution of C^α pair distances in proteins[43] in the hierarchical method,

or some knowledge of the position of the remnant(s) for the decimation method.

Conclusions

Several loop-closure problems for proteins have been solved exactly by reducing them to extracting the roots of a polynomial. For the tripeptide loop-closure problem, the degree of the polynomial is 16; hence, there are at most 16 real roots, and at most 16 loop closures for a given tripeptide problem, consistent with the empirically observed limit of eight loop-closure solutions.[1]

Both of these values are contradicted by a recent article claiming that only four solutions are possible.[44] That article did correctly show that no more than four solutions are possible for a fixed value of Ω_1 (a result originally derived by Go and Scheraga[1]). However, the authors overlooked the possibility that different values of Ω_1 may lead to valid loop closures, and furthermore, the bond angle constraint at C_3^α was not considered. Therefore, their results do not invalidate the results derived here or the observations of the original article.[1]

The coefficients of the loop-closure polynomial are real and, therefore, the number of real roots (and thus, the number of valid loop closures) must be even, except in exceptional cases where there is a double root. In particular, the existence of one valid loop closure implies the existence of another.

The general seven-atom loop-closure problem was shown to be equivalent to the nine-atom tripeptide loop-closure problem, and thus, also has at most 16 solutions. The eight-atom disulfide-bonded loop-closure problem was likewise shown to be equivalent to determining the real roots of a polynomial of degree 32. The Sturm chain method provides a method for counting the number of real roots of these polynomials without having to extract them. Several geometric feasibility criteria were also given to quickly determine whether a given tripeptide loop-closure problem has any solutions, and to set bounds on its solutions.

Three methods were offered for extending the tripeptide, seven-atom, and eight-atom loop-closure solutions to longer loops. The first was to sample all the dihedral angles of the loop except for the middle six, which may be determined from the tripeptide loop-closure algorithm. The sampling of the outer dihedral angles may be systematic (grid search) or stochastic, although the former method becomes prohibitively expensive in longer loops. The geometrical criteria for loop closure as well as the Sturm chain method provide rapid checks whether any loop closures are possible for a given choice of the outer dihedral angles.

The second method involves decimating the longer loop into rigid fragments called remnants. These remnants may be moved quasi-independently, and the intervening backbone atoms can be reconstituted by applying a loop-closure algorithm to join the remnants together again. Such a method is well suited when certain remnants are known to lie in a specific region.

The third method involves applying the loop closure algorithm hierarchically. For example, suppose that the distances and angles between every fourth C^α atom are sampled. Then the loop-closure algorithm can be used to reconstruct the positions of those C^α atoms; then the remaining backbone atoms can be reconstituted using the loop-closure algorithm a second time. This also has the advantage that statistical information about the distribution of distances between C^α atoms can be applied, which conveys much of the same information as secondary structure propensities.[43]

One noteworthy feature of the algorithms developed here is that they are purely geometrical, and take no account of atomic interactions, favorable or otherwise. In particular, no provision is made to avoid atomic overlaps; the algorithms generate all conformations consistent with the bond lengths and bond angles. However, atomic overlaps can be removed relatively quickly by energy minimization; a procedure to minimize the energies of closed loops has been presented elsewhere.[18]

This article has emphasized analytical methods, for example, resultants, to find all the loop-closure solutions, but numerical methods or a hybrid method may prove more efficient; further testing is warranted to find the method of least computational expense. It should also be noted that the methods described here are not immediately applicable to loops containing prolines, which effectively have only 1 degree of freedom, rather than 2.[12, 32, 33] However, it is not difficult to extend the methods to close loops containing prolines.

The main advantage of the loop-closure algorithms lies in Monte Carlo sampling, because they allow rearrangements as the elementary steps in a Monte Carlo simulation. This localization of movement brings two benefits. First, energy evaluations are much faster; a 10-fold reduction in the number of moving atoms results in (at least) a fivefold reduction in the computational cost of an energy

evaluation, and possibly much more if the individual fixed atoms are replaced with an equivalent field.[45] Second, the moves are less likely to destroy well-optimized substructures, unlike uncorrelated changes in the dihedral angles. In this way, loop-closure algorithms for proteins are analogous to cluster algorithms for Ising models[46] and the efficient move-set developed by Lin[47] for the traveling salesman problem.

These algorithms may serve an important role in other protein studies. For example, hinge motions in domains can be investigated by sampling the relative positions of the two domains and using the loop-closure algorithm to reconnect them. Similarly, the tertiary structure of a protein could be predicted from its secondary structure by holding the secondary structure elements rigid, sampling their packing, and using the loop-closure algorithms to connect the elements.

Loop-closure algorithms may also be applied to the NMR problems, for example, the determination of the packing of rigid subunits. Interatomic distances can be estimated from NMR data using the NOE (nuclear Overhauser effect) method. If the NOE distances have the form of pseudobond lengths and pseudobond angles, loop-closure methods may be used to solve for the relative position of one subunit with respect to the other (Fig. 20).

Loop-closure algorithms may also be used to regularize the geometry of a PDB structure. Molecular simulations with rigid geometry potentials such as ECEPP[33] require that the covalent bond geometry be standardized to that with which the force field was parametrized. This standardization can be accomplished readily by cutting out short nonstandard loops and rebuilding standard geometry loops in their place. By repeating this procedure with overlapping loops, the whole protein can easily be brought to standard geometry.

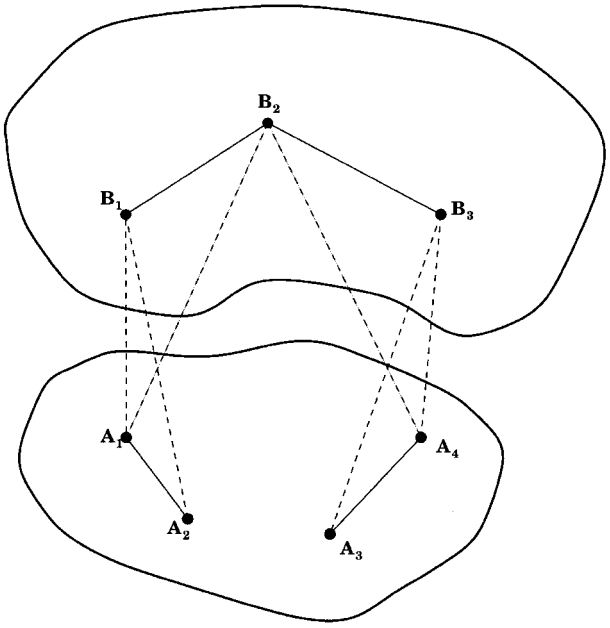

FIGURE 20. The relative positions of two rigid subunits A and B can be found if the six intersubunit distance constraints have the form of a loop closure, i.e., of rigid ''bond'' lengths and ''bond'' angles, as in the dashed lines of this figure. Such distance constraints can be derived from experiment, for example, NOESY data. The bond lengths and bond angles within each rigid subunit are, of course, fixed. The positions of the atoms B_1, B_2, and B_3 can be found by using the general seven-atom loop-closure method described in the text; the positions of these three atoms suffice to fix the position and orientation of subunit B relative to subunit A.

Acknowledgments

We thank R. Connelly for helpful discussions of Bezout's theorem, and the Cornell Center for Materials Research for computer resources to test the algorithm and make the figures. The Cornell Center for Materials Research is supported by the National Science Foundation (NSF Funded Award DMR-9632275). This work was supported by grants from the National Science Foundation (MCB95-13167) and from the National Institutes of Health (GM-14312).

Appendix A: Deriving the First Biquadratic Equation

This Appendix solves for the relation between the two spherical angles Ω_1 and ζ_1 in the first spherical quadrangle (Fig. 6), and derives the first of three biquadratic equations that will be used to solve the tripeptide loop-closure problem. The derivation is carried out here in detail, and is referenced in the subsequent derivations of the second and third biquadratic equations (Appendices B and C).

The geometric variables are defined as follows. Let there be four points on the unit sphere labeled $\hat{\mathbf{p}}_1$, $\hat{\mathbf{r}}$, $\hat{\mathbf{q}}_1$, and $\hat{\mathbf{s}}_1$, and let them be connected in a quadrangle by geodesic arcs of fixed length α_1, β_1, γ_1, and δ_1, respectively (Figs. 4 and 6). Let $\hat{\mathbf{p}}_1$ and $\hat{\mathbf{q}}_1$ be joined by a geodesic diagonal of length λ_1, and let the opposite spherical angle $\angle\hat{\mathbf{p}}_1\hat{\mathbf{r}}\hat{\mathbf{q}}_1$ be

denoted by Ω_1. Furthermore, let the spherical angle $\angle\hat{\mathbf{r}}\hat{\mathbf{q}}_1\hat{\mathbf{s}}_1$ be denoted by ζ_1, which is divided by the diagonal λ_1 into ζ_A and ζ_B, defined by the angles $\angle\hat{\mathbf{r}}\hat{\mathbf{q}}_1\hat{\mathbf{p}}_1$ and $\angle\hat{\mathbf{p}}_1\hat{\mathbf{q}}_1\hat{\mathbf{s}}_1$, respectively.

We require an equation relating the spherical angle ζ_1 to the angle Ω_1 for a given set of fixed lengths α_1, β_1, γ_1, and ζ_1. This fundamental equation will be derived from the following relation, which results from the law of cosines for spherical trigonometry[35]

$$\cos\delta_1 = \cos\gamma_1\cos\lambda_1 + \sin\gamma_1\sin\lambda_1\cos\zeta_B \quad \text{(A1)}$$

The variables λ_1 and ζ_B will be eliminated in favor of the fixed quantities α_1, β_1, γ_1, δ_1, and the variables Ω_1 and ζ_1.

The $\cos\lambda_1$ term on the right-hand side of eq. (A1) is relatively easy to eliminate, using the law of cosines

$$\cos\lambda_1 = \cos\alpha_1\cos\beta_1 + \sin\alpha_1\sin\beta_1\cos\Omega_1 \quad \text{(A2)}$$

However, the $\sin\lambda_1\cos\zeta_B$ term is more difficult, and requires consideration of the additional spherical angle ζ_A. The spherical angles ζ_1, ζ_A, and ζ_B are related by a trigonometric identity

$$\begin{aligned}\cos\zeta_B &= \cos(\zeta_1 - \zeta_A)\\ &= \cos\zeta_1\cos\zeta_A + \sin\zeta_1\sin\zeta_A \quad \text{(A3)}\end{aligned}$$

Multiplying by $\sin\lambda_1$ yields the equation

$$\sin\lambda_1\cos\zeta_B = \sin\lambda_1(\cos\zeta_1\cos\zeta_A + \sin\zeta_1\sin\zeta_A) \quad \text{(A4)}$$

$$\begin{aligned}&= \cos\zeta_1\sin\lambda_1\cos\zeta_A\\ &\quad + \sin\zeta_1\sin\lambda_1\sin\zeta_A \quad \text{(A5)}\end{aligned}$$

The $\sin\lambda_1\sin\zeta_A$ term on the right-hand side of eq. (A5) can be eliminated using the law of sines for spherical trigonometry[35]

$$\sin\lambda_1\sin\zeta_A = \sin\alpha_1\sin\Omega_1 \quad \text{(A6)}$$

The $\sin\lambda_1\cos\zeta_A$ term can also be eliminated, albeit with more work. The law of cosines requires that

$$\sin\beta_1\sin\lambda_1\cos\zeta_A = \cos\alpha_1 - \cos\beta_1\cos\lambda_1 \quad \text{(A7)}$$

Substituting eq. (A2) for $\cos\lambda_1$ in equation (A7) gives the result

$$\begin{aligned}&\sin\beta_1\sin\lambda_1\cos\zeta_A\\ &\quad = \cos\alpha_1(1 - \cos^2\beta_1)\\ &\qquad - \sin\alpha_1\sin\beta_1\cos\beta_1\cos\Omega_1 \quad \text{(A8)}\\ &\quad = \cos\alpha_1\sin^2\beta_1 - \sin\alpha_1\sin\beta_1\cos\beta_1\cos\Omega_1 \quad \text{(A9)}\end{aligned}$$

Dividing both sides by $\sin\beta_1$ yields the needed relation

$$\sin\lambda_1\cos\zeta_A = \cos\alpha_1\sin\beta_1 - \sin\alpha_1\cos\beta_1\cos\Omega_1 \quad \text{(A10)}$$

Using equations (A6) and (A10), the $\sin\lambda_1\sin\zeta_A$ and $\sin\lambda_1\cos\zeta_A$ terms can be eliminated from eq. (A5) in favor of the given quantities α_1 and β_1 and the variables Ω_1 and ζ_1

$$\begin{aligned}&\sin\lambda_1\cos\zeta_B\\ &\quad = \cos\zeta_1(\cos\alpha_1\sin\beta_1 - \sin\alpha_1\cos\beta_1\cos\Omega_1)\\ &\qquad + \sin\zeta_1\sin\Omega_1\sin\alpha_1 \quad \text{(A11)}\end{aligned}$$

Finally, the fundamental equation (A1) assumes the desired form upon substituting eq. (A2) for $\cos\lambda_1$ and eq. (A11) for $\sin\lambda_1\cos\zeta_B$

$$\begin{aligned}\cos\delta_1 &= \cos\gamma_1(\cos\alpha_1\cos\beta_1\\ &\qquad + \sin\alpha_1\sin\beta_1\cos\Omega_1)\\ &\qquad + \sin\gamma_1\cos\zeta_1(\cos\alpha_1\sin\beta_1\\ &\qquad - \sin\alpha_1\cos\beta_1\cos\Omega_1)\\ &\qquad + \sin\gamma_1\sin\zeta_1\sin\Omega_1\sin\alpha_1 \quad \text{(A12)}\end{aligned}$$

Collecting the fixed quantities into constants yields the simple equation

$$C_0 + C_1\cos\Omega_1 + C_2\cos\zeta_1 + C_3\cos\Omega_1\cos\zeta_1 + C_4\sin\Omega_1\sin\zeta_1 = 0 \quad \text{(A13)}$$

where the constants are defined by

$$C_0 \equiv \cos\alpha_1\cos\beta_1\cos\gamma_1 - \cos\delta_1 \quad \text{(A14)}$$

$$C_1 \equiv \sin\alpha_1\sin\beta_1\cos\gamma_1 \quad \text{(A15)}$$

$$C_2 \equiv \cos\alpha_1\sin\beta_1\sin\gamma_1 \quad \text{(A16)}$$

$$C_3 \equiv -\sin\alpha_1\cos\beta_1\sin\gamma_1 \quad \text{(A17)}$$

$$C_4 \equiv \sin\alpha_1\sin\gamma_1 \quad \text{(A18)}$$

The fundamental equation (A13) can also be recast as a polynomial equation using the trigonometric identities

$$\cos\zeta_1 = \frac{1 - z_1^2}{1 + z_1^2} \quad \sin\zeta_1 = \frac{2z_1}{1 + z_1^2} \quad \text{(A19)}$$

$$\cos\Omega_1 = \frac{1 - w_1^2}{1 + w_1^2} \quad \sin\Omega_1 = \frac{2w_1}{1 + w_1^2} \quad \text{(A20)}$$

where z_1 and w_1 are defined by the half-angle formulae

$$z_1 \equiv \tan\left(\frac{\zeta_1}{2}\right) \quad \text{(A21)}$$

$$w_1 \equiv \tan\left(\frac{\Omega_1}{2}\right) \quad \text{(A22)}$$

Substituting eqs. (A19) and (A20) into eq. (A13), and multiplying by $(1 + w_1^2)(1 + z_1^2)$ yields the polynomial form of the fundamental equation

$$D_0 + D_1 w_1^2 + D_2 z_1^2 + D_3 w_1^2 z_1^2 + D_4 w_1 z_1 = 0 \quad \text{(A23)}$$

whose coefficients are given by

$$D_0 \equiv C_0 + C_1 + C_2 + C_3 \quad \text{(A24)}$$

$$D_1 \equiv C_0 - C_1 + C_2 - C_3 \quad \text{(A25)}$$

$$D_2 \equiv C_0 + C_1 - C_2 - C_3 \quad \text{(A26)}$$

$$D_3 \equiv C_0 - C_1 - C_2 + C_3 \quad \text{(A27)}$$

$$D_4 \equiv 4C_4 \quad \text{(A28)}$$

This general solution for the spherical quadrangle will be used to formulate biquadratic equations for the two other quadrangles mentioned above.

Appendix B: Deriving the Second Biquadratic Equation

The second quadrangle is defined by the four unit vectors $\hat{\mathbf{s}}_3$, $\hat{\mathbf{r}}$, $\hat{\mathbf{q}}_3$, and $\hat{\mathbf{p}}_3$ (Fig. 7). Using the general solution of Appendix A, a relationship can be established between the spherical angle ζ_3 (defined as the angle $\angle\hat{\mathbf{r}}\hat{\mathbf{q}}_3\hat{\mathbf{p}}_3$) and the angle Ω_3 (defined as the angle $\angle\hat{\mathbf{s}}_3\hat{\mathbf{r}}\hat{\mathbf{q}}_3$); by analogy with eq. (A13), we obtain the equation

$$N_0 + N_1 \cos\Omega_3 + N_2 \cos\zeta_3 + N_3 \cos\Omega_3 \cos\zeta_3 + N_4 \sin\Omega_3 \sin\zeta_3 = 0 \quad \text{(B1)}$$

where the constants $N_0 - N_4$ are given by

$$N_0 \equiv \cos\alpha_3 \cos\beta_3 \cos\gamma_3 - \cos\delta_3 \quad \text{(B2)}$$

$$N_1 \equiv \sin\alpha_3 \sin\beta_3 \cos\gamma_3 \quad \text{(B3)}$$

$$N_2 \equiv \cos\alpha_3 \sin\beta_3 \sin\gamma_3 \quad \text{(B4)}$$

$$N_3 \equiv -\sin\alpha_3 \cos\beta_3 \sin\gamma_3 \quad \text{(B5)}$$

$$N_4 \equiv \sin\alpha_3 \sin\gamma_3 \quad \text{(B6)}$$

However, it is preferable to eliminate Ω_3 in favor of Ω_1, which is already a variable in the first biquadratic equation. By our geometric definitions (Appendix A),their relationship is

$$\Omega_3 + \Omega_0 = \Omega_1 \quad \text{(B7)}$$

where Ω_0 is the given spherical angle $\angle\hat{\mathbf{p}}_1\hat{\mathbf{r}}\hat{\mathbf{s}}_3$. Thus, it follows from a trigonometric identity that

$$\cos\Omega_3 = \cos\Omega_1 \cos\Omega_0 + \sin\Omega_1 \sin\Omega_0 \quad \text{(B8)}$$

$$\sin\Omega_3 = \sin\Omega_1 \cos\Omega_0 - \cos\Omega_1 \sin\Omega_0 \quad \text{(B9)}$$

Substituting of eqs. (B8) and (B9) into eq. (B1) yields an equation of eight terms

$$N_0 + N_1 \cos\Omega_0 \cos\Omega_1 + N_1 \sin\Omega_0 \sin\Omega_1 + N_2 \cos\zeta_3 + N_3 \cos\Omega_0 \cos\Omega_1 \cos\zeta_3 + N_3 \sin\Omega_0 \sin\Omega_1 \cos\zeta_3 + N_4 \cos\Omega_0 \sin\Omega_1 \sin\zeta_3 - N_4 \sin\Omega_0 \cos\Omega_1 \sin\zeta_3 = 0 \quad \text{(B10)}$$

This equation may be converted into a polynomial equation using the trigonometric identities of Appendix A

$$\cos\zeta_3 = \frac{1 - z_3^2}{1 + z_3^2} \qquad \sin\zeta_3 = \frac{2z_3}{1 + z_3^2} \quad \text{(B11)}$$

$$\cos\Omega_1 = \frac{1 - w_1^2}{1 - w_1^2} \qquad \sin\Omega_1 = \frac{2w_1}{1 + w_1^2} \quad \text{(B12)}$$

where w_1 and z_3 are the tangents of the half-angles of Ω_1 and ζ_3, respectively. Substituting these equations and multiplying $(1 + w_1^2)(1 + z_3^2)$ yields the second biquadratic equation

$$P_{00} + P_{10} w_1 + P_{01} z_3 + P_{11} w_1 z_3 + P_{20} w_1^2 + P_{02} z_3^2 + P_{12} w_1 z_3^2 + P_{21} w_1^2 z_3 + P_{22} w_1^2 z_3^2 = 0 \quad \text{(B13)}$$

whose coefficients are given by

$$P_{00} \equiv N_0 + N_1 \cos\Omega_0 + N_2 + N_3 \cos\Omega_0 \quad \text{(B14)}$$

$$P_{10} \equiv 2(N_1 + N_3)\sin\Omega_0 \quad \text{(B15)}$$

$$P_{01} \equiv -2N_4 \sin\Omega_0 \quad \text{(B16)}$$

$$P_{11} \equiv 4N_4 \cos\Omega_0 \quad \text{(B17)}$$

$$P_{20} \equiv N_0 - N_1 \cos\Omega_0 + N_2 - N_3 \cos\Omega_0 \quad \text{(B18)}$$

$$P_{02} \equiv N_0 + N_1 \cos\Omega_0 - N_2 - N_3 \cos\Omega_0 \quad \text{(B19)}$$

$$P_{12} \equiv 2(N_1 - N_3)\sin\Omega_0 \quad \text{(B20)}$$

$$P_{21} \equiv 2N_4 \sin\Omega_0 \quad \text{(B21)}$$

$$P_{22} \equiv N_0 - N_1 \cos\Omega_0 - N_2 + N_3 \cos\Omega_0 \quad \text{(B22)}$$

Appendix C: Deriving the Third Biquadratic Equation

The third quadrangle is defined by the four unit vectors $\hat{\mathbf{p}}_2$, $\hat{\mathbf{q}}_1$, $\hat{\mathbf{q}}_3$, and $\hat{\mathbf{s}}_2$, as depicted in Figure 8. The planarity of the peptide group ensures that $\hat{\mathbf{q}}_1$, $\hat{\mathbf{p}}_2$, and $\hat{\mathbf{s}}_1$ are coplanar, and thus lie on the same great circle; hence, the spherical angle $\angle\hat{\mathbf{r}}\hat{\mathbf{q}}_1\hat{\mathbf{p}}_2$ equals the angle $\angle\hat{\mathbf{r}}\hat{\mathbf{q}}_1\hat{\mathbf{s}}_1$, which has been labeled ζ_1 in Figure 5. For analogous reasons, the spherical angle $\angle\hat{\mathbf{r}}\hat{\mathbf{q}}_3\hat{\mathbf{s}}_2$ equals the angle $\angle\hat{\mathbf{r}}\hat{\mathbf{q}}_3\hat{\mathbf{p}}_3$, which has been labeled ζ_3 in Figure 5. Last, because $\hat{\mathbf{r}}$, $\hat{\mathbf{q}}_1$, and $\hat{\mathbf{q}}_3$ are coplanar by construction, the angles ζ_1 and ζ_3 also refer to the supplement of the spherical angle $\angle\hat{\mathbf{p}}_2\hat{\mathbf{q}}_1\hat{\mathbf{q}}_3$ and to the spherical angle $\angle\hat{\mathbf{q}}_1\hat{\mathbf{q}}_3\hat{\mathbf{s}}_2$, respectively (Fig. 8).

The interior spherical angle η_1 is the supplement of the angle ζ_1, and thus the following relations hold

$$\sin\eta_1 = \sin(\pi - \zeta_1) = \sin\zeta_1 \tag{C1}$$

$$\cos\eta_1 = \cos(\pi - \zeta_1) = -\cos\zeta_1 \tag{C2}$$

The angles ζ_3 and η_1 are related by the solution given in Appendix A

$$L_0 + L_1\cos\eta_1 + L_2\cos\zeta_3 + L_3\cos\eta_1\cos\zeta_3 + L_4\sin\eta_1\sin\zeta_3 = 0 \tag{C3}$$

where

$$L_0 \equiv \cos\alpha_2\cos\beta_2\cos\gamma_2 - \cos\delta_2 \tag{C4}$$

$$L_1 \equiv \sin\alpha_2\sin\beta_2\cos\gamma_2 \tag{C5}$$

$$L_2 \equiv \cos\alpha_2\sin\beta_2\sin\gamma_2 \tag{C6}$$

$$L_3 \equiv -\sin\alpha_2\cos\beta_2\sin\gamma_2 \tag{C7}$$

$$L_4 \equiv \sin\alpha_2\sin\gamma_2 \tag{C8}$$

Substituting equations (C1) and (C2) for $\sin\eta_1$ and $\cos\eta_1$ yields

$$L_0 - L_1\cos\zeta_1 + L_2\cos\zeta_3 - L_3\cos\zeta_1\cos\zeta_3 + L_4\sin\zeta_1\sin\zeta_3 = 0 \tag{C9}$$

This equation can, likewise, be converted to a polynomial using the trigonometric identities of Appendix A

$$\cos\zeta_1 = \frac{1 - z_1^2}{1 + z_1^2} \quad \sin\zeta_1 = \frac{2z_1}{1 + z_1^2} \tag{C10}$$

$$\cos\zeta_3 = \frac{1 - z_3^2}{1 + z_3^2} \quad \sin\zeta_3 = \frac{2z_3}{1 + z_3^2} \tag{C11}$$

where z_1 and z_3 are the tangents of the half-angles of ζ_1 and ζ_3, respectively. Substituting these equations and multiplying by $(1 + z_1^2)(1 + z_3^2)$ yields the third biquadratic equation

$$M_0 + M_1 z_1^2 + M_2 z_3^2 + M_3 z_1^2 z_3^2 + M_4 z_1 z_3 = 0 \tag{C12}$$

whose coefficients are given by

$$M_0 \equiv L_0 - L_1 + L_2 - L_3 \tag{C13}$$

$$M_1 \equiv L_0 + L_1 + L_2 + L_3 \tag{C14}$$

$$M_2 \equiv L_0 - L_1 - L_2 + L_3 \tag{C15}$$

$$M_3 \equiv L_0 + L_1 - L_2 - L_3 \tag{C16}$$

$$M_4 \equiv 4L_4 \tag{C17}$$

Appendix D: Solving the Three Biquadratic Equations

The solution of the tripeptide loop closure problem has been converted into solving three coupled biquadratic equations in three variables w_1, z_1, and z_3. These equations may be solved by the method of resultants,[48] as described in this section.

The three biquadratic equations are given by

$$D_0 + D_1 w_1^2 + D_2 z_1^2 + D_3 w_1^2 z_1^2 + D_4 w_1 z_1 = 0 \tag{D1}$$

$$M_0 + M_1 z_1^2 + M_2 z_3^2 + M_3 z_1^2 z_3^2 + M_4 z_1 z_3 = 0 \tag{D2}$$

and

$$P_{00} + P_{10}w_1 + P_{01}z_3 + P_{11}w_1 z_3 + P_{20}w_1^2 + P_{02}z_3^2 + P_{12}w_1 z_3^2 + P_{21}w_1^2 z_3 + P_{22}w_1^2 z_3^2 = 0 \tag{D3}$$

To eliminate z_1, eqs. (D1) and (D2) may be written in square 4×4 matrix form

$$U \cdot \begin{pmatrix} 1 \\ z_1 \\ z_1^2 \\ z_1^3 \end{pmatrix} \equiv \begin{bmatrix} U_{11} & U_{12} & U_{13} & 0 \\ 0 & U_{11} & U_{12} & U_{13} \\ U_{31} & U_{32} & U_{33} & 0 \\ 0 & U_{31} & U_{32} & U_{33} \end{bmatrix} \begin{pmatrix} 1 \\ z_1 \\ z_1^2 \\ z_1^3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{D4}$$

where the matrix elements are given by

$$U_{11} \equiv D_0 + D_1 w_1^2 \quad \text{(D5)}$$
$$U_{12} \equiv D_1 w_1 \quad \text{(D6)}$$
$$U_{13} \equiv D_2 + D_3 w_1^2 \quad \text{(D7)}$$
$$U_{31} \equiv M_0 + M_2 z_3^2 \quad \text{(D8)}$$
$$U_{32} \equiv M_4 z_3 \quad \text{(D9)}$$
$$U_{33} \equiv M_1 + M_3 z_3^2 \quad \text{(D10)}$$

It should be noted that the matrix **U** depends only on the variables w_1 and z_3, and not on the variable z_1. Following the laws of linear algebra, the det(**U**) must be zero, which conveniently eliminates z_1 and gives a new equation

$$\begin{aligned} &Q_{00} + Q_{20} w_1^2 + Q_{40} w_1^4 + Q_{11} w_1 z_3 + Q_{31} w_1^3 z_3 \\ &\quad + Q_{02} z_3^2 + Q_{22} w_1^2 z_3^2 + Q_{42} w_1^4 z_3^2 + Q_{13} w_1 z_3^3 \\ &\quad + Q_{33} w_1^3 z_3^3 + Q_{04} z_3^4 + Q_{24} w_1^2 z_3^4 + Q_{44} w_1^4 z_3^4 \\ &\quad = 0 \end{aligned} \quad \text{(D11)}$$

where the 13 coefficients are defined by

$$Q_{00} \equiv (D_0 M_1 - D_2 M_0)^2 \quad \text{(D12)}$$
$$Q_{20} \equiv 2(D_0 M_1 - D_2 M_0)(D_1 M_1 - D_3 M_0) + D_4^2 M_0 M_1 \quad \text{(D13)}$$
$$Q_{40} \equiv (D_1 M_1 - D_3 M_0)^2 \quad \text{(D14)}$$
$$Q_{11} \equiv -D_4 M_4 (D_0 M_1 + D_2 M_0) \quad \text{(D15)}$$
$$Q_{31} \equiv -D_4 M_4 (D_1 M_1 + D_3 M_0) \quad \text{(D16)}$$
$$Q_{02} \equiv 2 D_0^2 M_1 M_3 + D_0 D_2 (M_4^2 - 2 M_0 M_3 - 2 M_1 M_2) + 2 D_2^2 M_0 M_2 \quad \text{(D17)}$$
$$Q_{22} \equiv 4 D_0 D_1 M_1 M_3 + (D_0 D_3 + D_1 D_2) \times (M_4^2 - 2 M_0 M_3 - 2 M_1 M_2) + 4 D_2 D_3 M_0 M_2 + D_4^2 (M_0 M_3 + M_1 M_2) \quad \text{(D18)}$$
$$Q_{42} \equiv 2 D_1^2 M_1 M_3 + D_1 D_3 (M_4^2 - 2 M_0 M_3 - 2 M_1 M_2) + 2 D_3^2 M_0 M_2 \quad \text{(D19)}$$
$$Q_{13} \equiv -D_4 M_4 (D_0 M_3 + D_2 M_2) \quad \text{(D20)}$$
$$Q_{33} \equiv -D_4 M_4 (D_1 M_3 + D_3 M_2) \quad \text{(D21)}$$
$$Q_{04} \equiv (D_0 M_3 - D_2 M_2)^2 \quad \text{(D22)}$$
$$Q_{24} \equiv 2(D_0 M_3 - D_2 M_2)(D_1 M_3 - D_3 M_2) + D_4^2 M_2 M_3 \quad \text{(D23)}$$
$$Q_{44} \equiv (D_1 M_3 - D_3 M_2)^2 \quad \text{(D24)}$$

The theory of resultants[48] shows that the two equations (D1) and (D2) have a valid solution for z_1 if and only if equation (D11) is satisfied. The value for z_1, which satisfies both equations, can be found by multiplying the first row of matrix **U** by U_{33} and the third row by U_{13} and subtracting, giving the equation

$$z_1 = \frac{U_{31} U_{13} - U_{11} U_{33}}{U_{12} U_{33} - U_{13} U_{32}} \quad \text{(D25)}$$

Thus, a unique value for z_1 can be found given the values of w_1 and z_3.

The variables w_1 and z_3 satisfy eqs. (D3) and (D11). A polynomial equation in w_1 alone can be obtained by eliminating z_3 using the method of resultants, as before. The two equations can be written as a 6×6 square matrix

$$\mathbf{V} \cdot \begin{pmatrix} 1 \\ z_3 \\ z_3^2 \\ z_3^3 \\ z_3^4 \\ z_3^5 \end{pmatrix} \equiv \begin{bmatrix} A_0 & A_1 & A_2 & A_3 & A_4 & 0 \\ 0 & A_0 & A_1 & A_2 & A_3 & A_4 \\ B_0 & B_1 & B_2 & 0 & 0 & 0 \\ 0 & B_0 & B_1 & B_2 & 0 & 0 \\ 0 & 0 & B_0 & B_1 & B_2 & 0 \\ 0 & 0 & 0 & B_0 & B_1 & B_2 \end{bmatrix} \begin{pmatrix} 1 \\ z_3 \\ z_3^2 \\ z_3^3 \\ z_3^4 \\ z_3^5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{(D26)}$$

where the matrix elements are given by

$$A_0 \equiv Q_{00} + Q_{20} w_1^2 + Q_{40} w_1^4 \quad \text{(D27)}$$
$$A_1 \equiv Q_{11} w_1 + Q_{31} w_1^3 \quad \text{(D28)}$$
$$A_2 \equiv Q_{02} + Q_{22} w_1^2 + Q_{42} w_1^4 \quad \text{(D29)}$$
$$A_3 \equiv Q_{13} w_1 + Q_{33} w_1^3 \quad \text{(D30)}$$
$$A_4 \equiv Q_{04} + Q_{24} w_1^2 + Q_{44} w_1^4 \quad \text{(D31)}$$
$$B_0 \equiv P_{00} + P_{10} w_1 + P_{20} w_1^2 \quad \text{(D32)}$$
$$B_1 \equiv P_{01} + P_{11} w_1 + P_{21} w_1^2 \quad \text{(D33)}$$
$$B_2 \equiv P_{02} + P_{12} w_1 + P_{22} w_1^2 \quad \text{(D34)}$$

The determinant of **V** must be zero, giving the relation

$$\begin{aligned} & A_0^2 B_2^4 + 2 A_0 A_4 B_0^2 B_2^2 - 4 A_0 A_4 B_0 B_1^2 B_2 \\ & \quad - 2 A_0 A_2 B_0 B_2^3 + 3 A_0 A_3 B_0 B_1 B_2^2 \\ & \quad - A_0 A_3 B_1^3 B_2 + A_0 A_4 B_1^4 - A_0 A_1 B_1 B_2^3 \\ & \quad + A_0 A_2 B_1^2 B_2^2 + A_4^2 B_0^4 - A_3 A_4 B_0^3 B_1 \\ & \quad - 2 A_2 A_4 B_0^3 B_2 + A_3^2 B_0^3 B_2 - A_2 A_3 B_0^2 B_1 B_2 \\ & \quad + 3 A_1 A_4 B_0^2 B_1 B_2 + A_2 A_4 B_0^2 B_1^2 \\ & \quad + A_2^2 B_0^2 B_2^2 - 2 A_1 A_3 B_0^2 B_2^2 + A_1 A_3 B_0 B_1^2 B_2 \\ & \quad - A_1 A_2 B_0 B_1 B_2^2 + A_1^2 B_0 B_2^3 - A_1 A_4 B_0 B_1^3 = 0 \end{aligned} \tag{D35}$$

Upon substitution of the definitions of the matrix elements of **V**, this yields a polynomial equation of degree 16 in the single variable w_1

$$\begin{aligned} & R_0 + R_1 w_1 + R_2 w_1^2 + R_3 w_1^3 + R_4 w_1^4 \\ & \quad + R_5 w_1^5 + R_6 w_1^6 + R_7 w_1^7 + R_8 w_1^8 + R_9 w_1^9 \\ & \quad + R_{10} w_1^{10} + R_{11} w_1^{11} + R_{12} w_1^{12} + R_{13} w_1^{13} \\ & \quad + R_{14} w_1^{14} + R_{15} w_1^{15} + R_{16} w_1^{16} = 0 \end{aligned} \tag{D36}$$

From the fundamental theorem of algebra, this polynomial has 16 roots, of which the real roots correspond to valid loop closures. These roots may be easily found by using publicly available software;[5] fast algorithms exist for the extraction of the roots of a polynomial. The coefficients R_0–R_{16} of this polynomial were computed explicitly by the software package Maple,[49] and are given in the supplement to this article, but their form is rather lengthy. It is easier to compute them numerically, which is done in a computer program that is also included in the supplement. The computation of loop coefficients generally takes less than a millisecond of CPU time.

The theory of resultants requires that the polynomial equation (D36) must hold in order that the two equations (D3) and (D11) have a solution for z_3. For each value of w_1 obtained from eq. (D36), the corresponding value for z_3 can be found from the formula

$$z_3 = \frac{K_3 B_0 - A_0 B_2^3}{K_2 B_2 - K_3 B_1} \tag{D37}$$

where the variables K are defined

$$K_0 \equiv A_2 B_2 - A_4 B_0 \tag{D38}$$

$$K_1 \equiv A_3 B_2 - A_4 B_1 \tag{D39}$$

$$K_2 \equiv A_1 B_2^2 - K_1 B_0 \tag{D40}$$

$$K_3 \equiv K_0 B_2 - K_1 B_1 \tag{D41}$$

This formula may be obtained as follows. By multiplying the first row of matrix **V** by B_2 and the fifth row by A_4 and subtracting, a new equation is produced in which the z_3^4 term has been eliminated. Similarly, the cubic (z_3^3) and quadratic (z_3^2) terms can be eliminated by the analogous crossmultiplication of the leading coefficients with the fourth and third rows of matrix **V**, respectively, until only a single power of z_3 remains, giving the solution (D37). In this way, all possible solutions (w_1, z_3) can be found, and thus the solution for z_1, as described by eq. (D25). With the values for w_1, z_1, and z_3 in hand, one can find the corresponding spherical angles Ω_1, ζ_1, and ζ_3, and thus solve the loop-closure problem, as described in Appendix E.

Appendix E: Finding the Atomic Positions from the Solutions of the Polynomial Equations

This section describes how to obtain the actual atomic positions from the values of w_1, z_1, and z_3, which satisfy the three equations (D1), (D2), and (D3).

The corresponding angles Ω_1, ζ_1, and ζ_3 can be found using the relations

$$w_1 \equiv \tan\left(\frac{\Omega_1}{2}\right) \tag{E1}$$

$$z_1 \equiv \tan\left(\frac{\zeta_1}{2}\right) \tag{E2}$$

$$z_3 \equiv \tan\left(\frac{\zeta_3}{2}\right) \tag{E3}$$

The unit vector $\hat{\mathbf{q}}_1$ can be computed from Ω_1, because Ω_1 is the dihedral angle between the $\hat{\mathbf{p}}_1\hat{\mathbf{r}}$ and $\hat{\mathbf{r}}\hat{\mathbf{q}}_1$ planes, and both $\hat{\mathbf{p}}_1$ and $\hat{\mathbf{r}}$ are fixed. The virtual bond vector $\mathbf{q}_1$ can be computed by multiplying $\hat{\mathbf{q}}_1$ by the virtual bond length q_1. The other virtual bond vector $\mathbf{q}_3$ (and thence its unit vector $\hat{\mathbf{q}}_3$) can be found using the relation $\mathbf{q}_3 = \mathbf{q}_1 - \mathbf{r}$. The position of C_2^α is determined by adding $\mathbf{q}_1$ to the (known) position of C_1^α.

The unit vector $\hat{\mathbf{q}}_1$ and the spherical angle ζ_1 determine the unit vector $\hat{\mathbf{s}}_1$, because ζ_1 is the dihedral angle between the $\hat{\mathbf{r}}\hat{\mathbf{q}}_1$ and $\hat{\mathbf{q}}_1\hat{\mathbf{s}}_1$ (Fig. 6). The dihedral angle unit vector $\hat{\mathbf{p}}_3$ can be obtained from $\hat{\mathbf{q}}_3$ and ζ_3 by an analogous method (Fig. 7).

Positioning the atoms is straightforward, given these unit vectors. The atoms C'_1 and N_3 can be found by multiplying the unit vectors $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{p}}_3$, respectively, by the corresponding bond length and adding the resulting vector to the positions of C_1^α and C_3^α, respectively. The two remaining atoms (N_2 and C'_2) can be found most quickly as linear combinations of the positions of their fellow atoms in their respective peptide groups. As an example, the four atoms of the peptide group (C_1^α, C'_1, N_2, and C_2^α) are coplanar but not collinear. Therefore, the three atoms of known position (C_1^α, C'_1, and C_2^α) form a two-dimensional reference frame, in which the N_2 atom has fixed coordinates, because the four atoms are fixed relative to each other in a peptide group. Therefore, the coordinates of the N_2 atom can be expressed as a linear combination of those of the other three atoms. By analogy, a similar method can be used to find the coordinates of the C'_2 atom as a linear combination of the coordinates of the other three atoms belonging to its peptide group.

Appendix F: Using Sturm Chains to Count Real Roots

The idea of the method of Sturm chains is to obtain a sequence of polynomials, $g_0(w)$, $g_1(w)$, $g_2(w), \ldots, g_m(w)$, as described below. The (integer) function $\Sigma(a)$ is defined as the number of sign changes that occur as the polynomials $g_0(a)$, $g_1(a)$, $g_2(a), \ldots, g_m(a)$ are successively evaluated; that is, if $g_i(a)$ and $g_{i+1}(a)$ have distinct signs, then $\Sigma(a)$ is incremented by one. If a and b are not themselves roots of $g_0(w)$, and if $g_0(w)$ has only simple (not multiple) roots, then $\Sigma(a) - \Sigma(b)$ equals the number of real roots of $g_0(w)$ in the interval $[a, b]$.

The Sturm sequence of polynomials is obtained by using the Euclidean algorithm for finding the greatest common divisor of two polynomials.[35] The polynomial $g_0(w)$ is given, and $g_1(w)$ is defined to be its derivative. The polynomial $g_2(w)$ is defined as the remainder when $g_1(w)$ is divided into $g_0(w)$, i.e.,

$$g_0(w) = q(w)g_1(w) + g_2(w) \qquad \text{(F1)}$$

where $q(w)$ is some quotient polynomial, which can be found by long division. Similarly, $g_3(w)$ is the remainder when $g_2(w)$ is divided into $g_1(w)$, $g_4(w)$ is the remainder when $g_3(w)$ is divided into $g_2(w)$, and so on. The sequence is finite, because the degree of $g_{i+1}(w)$ is at least one degree less than that of $g_i(w)$. The last member of the sequence is denoted by $g_m(w)$, and represents the greatest common divisor of $g_0(w)$ and $g_1(w)$. If $g_m(w)$ is not a simple constant, i.e., if $g_0(w)$ and $g_1(w)$ have a common polynomial factor, then one must divide $g_0(w)$ by the common factor $g_m(w)$ and repeat the Euclidean algorithm to obtain a valid Sturm sequence, which requires that $g_m(w)$ be a simple constant. If this condition holds, the roots of $g_0(w)$ are all simple (not multiple) roots and the number of such simple roots in the interval $[a, b]$ is given by $\Sigma(a) - \Sigma(b)$.

For example, suppose that the polynomial $g_0(w)$ has the form[35]

$$g_0(w) = w^4 - 5w^2 + 8w - 8 \qquad \text{(F2)}$$

The next member in the Sturm sequence is the derivative

$$g_1(w) \equiv g'_0(w) = 4w^3 - 10w + 8 \qquad \text{(F3)}$$

Long division of $g_1(w)$ into $g_0(w)$ yields the remainder

$$g_2(w) = 5w^2 - 12w + 16 \qquad \text{(F4)}$$

where we premultiplied $g_0(w)$ by -2 to make the coefficients simple; this is permissible because we are concerned only with the signs of the Sturm sequence polynomials. Long division of $g_2(w)$ into $g_1(w)$ likewise yields the remainder

$$g_3(w) = -3x + 284 \qquad \text{(F5)}$$

and, similarly, long division of $g_3(w)$ into $g_2(w)$ yields the constant

$$g_4(w) = -1 \qquad \text{(F6)}$$

which terminates the Sturm sequence. Because -1 is a simple constant, $g_0(w)$ and $g_1(w)$ have no common polynomial factor, and the number of real roots in an interval equals the difference in Σ at its two end points.

Suppose that one wishes to compute the number of real positive roots of the polynomial (F2), i.e., the number of real roots between zero and positive infinity. Evaluating the chain at $w = 0$ yields the sequence $-8, 8, 16, 284, -1$, which has two sign changes; thus, $\Sigma(0)$ equals 2. At positive infinity, the signs are determined by the leading coefficients in the sequence; because the signs of the leading coefficients $(1, 4, 5, -3, -1)$ change sign once, $\Sigma(\infty)$ equals one. Taking the difference $\Sigma(0)$

$-\Sigma(\infty) = 1$ shows that there is one real, positive root for this polynomial. Similarly, the number of all real roots is given by the difference $\Sigma(-\infty) - \Sigma(\infty)$. In evaluating $\Sigma(-\infty)$, only the leading coefficients of the Sturm chain polynomials are significant, but they must be multiplied by -1 to the power of their leading exponent; the example above yields the sequence $1, -4, 5, 3, -1$, which has three sign changes, thus, $\Sigma(-\infty) = 3$. Taking the difference $\Sigma(-\infty) - \Sigma(\infty) = 2$ shows that there are two real roots for this polynomial.

Further details on Sturm sequences can be found in ref. 36.

Appendix G: Closing Disulfide-Bonded Loop

This Appendix provides the details of the calculation for closing the disulfide-bonded loop (Fig. 11).[37,38] The strategy of the method is to derive four equations relating the four spherical angles Ω_A, Ω_B, ζ_A, and ζ_B, which may be converted into four biquadratic equations by the change of variables

$$w_A \equiv \tan\left(\frac{\Omega_A}{2}\right) \tag{G1}$$

$$w_B \equiv \tan\left(\frac{\Omega_B}{2}\right) \tag{G2}$$

$$z_A \equiv \tan\left(\frac{\zeta_A}{2}\right) \tag{G3}$$

$$z_B \equiv \tan\left(\frac{\zeta_B}{2}\right) \tag{G4}$$

The four biquadratic equations in four variables can be reduced to an equation in a single variable by the method of resultants.

THE FIRST EQUATION

The first equation relating the spherical angles Ω_A and Ω_B may be derived as follows. The central quadrangle (Fig. 14) is equivalent to the vector equation

$$\mathbf{r}_2 = \mathbf{r}_1 - \mathbf{q}_A + \mathbf{q}_B, \tag{G5}$$

which, when squared, gives the equation

$$r_2^2 = r_1^2 + q_A^2 + q_B^2 - 2r_1q_A(\hat{\mathbf{r}}_1 \cdot \hat{\mathbf{q}}_A) + 2r_1q_B(\hat{\mathbf{r}}_1 \cdot \hat{\mathbf{q}}_B) - 2q_Aq_B(\hat{\mathbf{q}}_A \cdot \hat{\mathbf{q}}_B) \tag{G6}$$

where the vectors have been separated into their magnitudes (r_2, r_2, q_A, and q_B) and their unit vectors ($\hat{\mathbf{r}}_1$, $\hat{\mathbf{q}}_A$, and $\hat{\mathbf{q}}_B$). The unit vector $\hat{\mathbf{r}}_1$ is fixed, and the other two unit vectors may be expressed in terms of the fixed unit vectors $\hat{\mathbf{p}}_A$ and $\hat{\mathbf{p}}_A$ and of the spherical angles Ω_A and Ω_B.

To accomplish this, we introduce a coordinate frame at the atom C_A^α consisting of the (known) vectors $\hat{\mathbf{p}}_A$, $\hat{\mathbf{u}}_A$, and $\hat{\mathbf{v}}_A$, where the latter two vectors are defined by

$$\hat{\mathbf{u}}_A = [\hat{\mathbf{r}}_1 - \hat{\mathbf{p}}_A \cos\alpha_A]\left(\frac{1}{\sin\alpha_A}\right) \tag{G7}$$

$$\hat{\mathbf{v}}_A = \hat{\mathbf{p}}_A \times \hat{\mathbf{u}}_A \tag{G8}$$

In this basis, the unit vector $\hat{\mathbf{q}}_A$ has the form

$$\hat{\mathbf{q}}_A = \hat{\mathbf{p}}_A \cos\beta_A + \hat{\mathbf{u}}_A \sin\beta_A \cos\Omega_A + \hat{\mathbf{v}}_A \sin\beta_A \sin\Omega_A \tag{G9}$$

Similarly, a basis $\hat{\mathbf{p}}_B$, $\hat{\mathbf{u}}_B$, and $\hat{\mathbf{v}}_B$ can be established at the atom C_B^α. The unit vector $\hat{\mathbf{q}}_B$ can be expressed in terms of these (known) vectors

$$\hat{\mathbf{q}}_B = \hat{\mathbf{p}}_B \cos\beta_B + \hat{\mathbf{u}}_B \sin\beta_B \cos\Omega_B + \hat{\mathbf{v}}_B \sin\beta_B \sin\Omega_B \tag{G10}$$

Because the unit vectors $\hat{\mathbf{r}}_1$, $\hat{\mathbf{p}}_A$, $\hat{\mathbf{u}}_A$, $\hat{\mathbf{v}}_A$, $\hat{\mathbf{p}}_B$, $\hat{\mathbf{u}}_B$, and $\hat{\mathbf{v}}_B$ are known explicitly, the dot products can be computed

$$\hat{\mathbf{r}}_1 \cdot \hat{\mathbf{q}}_A \equiv \cos\xi_A = \cos\alpha_A \cos\beta_A + \sin\alpha_A \sin\beta_A \cos\Omega_A \tag{G11}$$

$$\hat{\mathbf{r}}_1 \cdot \hat{\mathbf{q}}_B \equiv \cos\xi_B = \cos\alpha_B \cos\beta_B + \sin\alpha_B \sin\beta_B \cos\Omega_B \tag{G12}$$

$$\begin{aligned}\hat{\mathbf{q}}_A \cdot \hat{\mathbf{q}}_B = {} & (\hat{\mathbf{p}}_A \cdot \hat{\mathbf{p}}_B)\cos\beta_A \cos\beta_B \\ & + (\hat{\mathbf{p}}_A \cdot \hat{\mathbf{u}}_B)\cos\beta_A \sin\beta_B \cos\Omega_B \\ & + (\hat{\mathbf{p}}_A \cdot \hat{\mathbf{v}}_B)\cos\beta_A \sin\beta_B \sin\Omega_B \\ & + (\hat{\mathbf{u}}_A \cdot \hat{\mathbf{p}}_B)\sin\beta_A \cos\beta_B \cos\Omega_A \\ & + (\hat{\mathbf{u}}_A \cdot \hat{\mathbf{u}}_B)\sin\beta_A \sin\beta_B \cos\Omega_A \cos\Omega_B \\ & + (\hat{\mathbf{u}}_A \cdot \hat{\mathbf{v}}_B)\sin\beta_A \sin\beta_B \cos\Omega_A \sin\Omega_B \\ & + (\hat{\mathbf{v}}_A \cdot \hat{\mathbf{p}}_B)\sin\beta_A \cos\beta_B \sin\Omega_A \\ & + (\hat{\mathbf{v}}_A \cdot \hat{\mathbf{u}}_B)\sin\beta_A \sin\beta_B \sin\Omega_A \cos\Omega_B \\ & + (\hat{\mathbf{v}}_A \cdot \hat{\mathbf{v}}_B)\sin\beta_A \sin\beta_B \sin\Omega_A \sin\Omega_B\end{aligned} \tag{G13}$$

Inserting these results into eq. (G6) yields the desired equation relating the spherical angles Ω_A and Ω_B

$$E_0 + E_1 \cos \Omega_A + E_2 \sin \Omega_A + E_3 \cos \Omega_B + E_4 \sin \Omega_B + E_5 \cos \Omega_A \cos \Omega_B + E_6 \cos \Omega_A \sin \Omega_B + E_7 \sin \Omega_A \cos \Omega_B + E_8 \sin \Omega_A \sin \Omega_B = 0 \quad \text{(G14)}$$

where the coefficients E_0–E_8 depend on the four fixed angles (α_A, α_B, β_A, β_B), the four fixed lengths (r_1, r_2, q_A, and q_B) and the nine fixed dot products between the basis vectors ($\hat{\mathbf{p}}_A$, $\hat{\mathbf{u}}_A$, $\hat{\mathbf{v}}_A$) and ($\hat{\mathbf{p}}_B$, $\hat{\mathbf{u}}_B$, $\hat{\mathbf{v}}_B$).

Using the trigonometric half-angle identities (G1) and (G2), this equation may be converted into a biquadratic equation in the variables w_A and w_B

$$F_0 + F_1 w_A + F_2 w_B + F_3 w_A w_B + F_4 w_A^2 + F_5 w_B^2 + F_6 w_A^2 w_B + F_7 w_A w_B^2 + F_8 w_A^2 w_B^2 = 0 \quad \text{(G15)}$$

where the nine coefficients F_0–F_8 are defined by the relations

$$F_0 \equiv E_0 + E_1 + E_3 + E_5 \quad \text{(G16)}$$
$$F_1 \equiv 2(E_2 + E_7) \quad \text{(G17)}$$
$$F_2 \equiv 2(E_4 + E_6) \quad \text{(G18)}$$
$$F_3 \equiv 4E_8 \quad \text{(G19)}$$
$$F_4 \equiv E_0 - E_1 + E_3 - E_5 \quad \text{(G20)}$$
$$F_5 \equiv E_0 + E_1 - E_3 - E_5 \quad \text{(G21)}$$
$$F_6 \equiv 2(E_4 - E_6) \quad \text{(G22)}$$
$$F_7 \equiv 2(E_2 - E_7) \quad \text{(G23)}$$
$$F_8 \equiv E_0 - E_1 - E_3 + E_5 \quad \text{(G24)}$$

This is the first biquadratic equation for eight-atom loops. An analogous equation was derived in an earlier article.[38]

THE SECOND EQUATION

The vector equation (G5) for the central quadrangle can also be written

$$\mathbf{r}_1 = \mathbf{r}_2 + \mathbf{q}_A - \mathbf{q}_B \quad \text{(G25)}$$

which, when squared, gives the equation

$$r_1^2 = r_2^2 + q_A^2 + q_B^2 + 2r_2 q_A(\hat{\mathbf{r}}_2 \cdot \hat{\mathbf{q}}_A) - 2r_2 q_B(\hat{\mathbf{r}}_2 \cdot \hat{\mathbf{q}}_B) - 2q_A q_B(\hat{\mathbf{q}}_A \cdot \hat{\mathbf{q}}_B) \quad \text{(G26)}$$

This equation will be used to derive the second biquadratic equation by expressing the dot products in terms of the spherical angles ζ_A and ζ_B.

To accomplish this, a coordinate frame $\hat{\mathbf{s}}_A$, $\hat{\mathbf{m}}_A$, and $\hat{\mathbf{n}}_A$ is introduced at the atom C_A^β where the latter two unit vectors are defined by

$$\hat{\mathbf{m}}_A \equiv [\hat{\mathbf{r}}_2 - \hat{\mathbf{s}}_A \cos \delta_A]\left(\frac{1}{\sin \delta_A}\right) \quad \text{(G27)}$$

$$\hat{\mathbf{n}}_A \equiv \hat{\mathbf{s}}_A \times \hat{\mathbf{m}}_A \quad \text{(G28)}$$

In this basis, the vector $\hat{\mathbf{q}}_A$ equals

$$\hat{\mathbf{q}}_A = \hat{\mathbf{s}}_A \cos \gamma_A + \hat{\mathbf{m}}_A \sin \gamma_A \cos \zeta_A + \hat{\mathbf{n}}_A \sin \gamma_A \sin \zeta_A \quad \text{(G29)}$$

Likewise, a coordinate frame, $\hat{\mathbf{s}}_B$ $\hat{\mathbf{m}}_B$, and $\hat{\mathbf{n}}_B$ can be introduced at the atom C_B^β where the latter two unit vectors are defined by

$$\hat{\mathbf{m}}_B \equiv [\hat{\mathbf{r}}_2 - \hat{\mathbf{s}}_B \cos \delta_B]\left(\frac{1}{\sin \delta_B}\right) \quad \text{(G30)}$$

$$\hat{\mathbf{n}}_B \equiv \hat{\mathbf{s}}_B \times \hat{\mathbf{m}}_B \quad \text{(G31)}$$

In this basis, the vector $\hat{\mathbf{q}}_B$ can be expressed

$$\hat{\mathbf{q}}_B = \hat{\mathbf{s}}_B \cos \gamma_B + \hat{\mathbf{m}}_B \sin \gamma_B \cos \zeta_B + \hat{\mathbf{n}}_B \sin \gamma_B \sin \zeta_B \quad \text{(G32)}$$

The vector $\hat{\mathbf{r}}_2$ can be expressed in the two frames by the relations

$$\hat{\mathbf{r}}_2 = \hat{\mathbf{s}}_A \cos \delta_A + \hat{\mathbf{m}}_A \sin \delta_A \quad \text{(G33)}$$
$$= \hat{\mathbf{s}}_B \cos \delta_B + \hat{\mathbf{m}}_B \sin \delta_B \quad \text{(G34)}$$

Therefore, the dot products are given by

$$\hat{\mathbf{r}}_2 \cdot \hat{\mathbf{q}}_A \equiv \cos \eta_A = \cos \gamma_A \cos \delta_A + \sin \gamma_A \sin \delta_A \cos \zeta_A \quad \text{(G35)}$$

$$\hat{\mathbf{r}}_2 \cdot \hat{\mathbf{q}}_B \equiv \cos \eta_B = \cos \gamma_B \cos \delta_B + \sin \gamma_B \sin \delta_B \cos \zeta_B \quad \text{(G36)}$$

$$\begin{aligned}\hat{\mathbf{q}}_A \cdot \hat{\mathbf{q}}_B = {} & (\hat{\mathbf{s}}_A \cdot \hat{\mathbf{s}}_B)\cos \gamma_A \cos \gamma_B \\ & + (\hat{\mathbf{s}}_A \cdot \hat{\mathbf{m}}_B)\cos \gamma_A \sin \gamma_B \cos \zeta_B \\ & + (\hat{\mathbf{s}}_A \cdot \hat{\mathbf{n}}_B)\cos \gamma_A \sin \gamma_B \sin \zeta_B \\ & + (\hat{\mathbf{m}}_A \cdot \hat{\mathbf{s}}_B)\sin \gamma_A \cos \gamma_B \cos \zeta_A \\ & + (\hat{\mathbf{m}}_A \cdot \hat{\mathbf{m}}_B)\sin \gamma_A \sin \gamma_B \cos \zeta_A \cos \zeta_B \\ & + (\hat{\mathbf{m}}_A \cdot \hat{\mathbf{n}}_B)\sin \gamma_A \sin \gamma_B \cos \zeta_A \sin \zeta_B \\ & + (\hat{\mathbf{n}}_A \cdot \hat{\mathbf{s}}_B)\sin \gamma_A \cos \gamma_B \sin \zeta_A \\ & + (\hat{\mathbf{n}}_A \cdot \hat{\mathbf{m}}_B)\sin \gamma_A \sin \gamma_B \sin \zeta_A \cos \zeta_B \\ & + (\hat{\mathbf{n}}_A \cdot \hat{\mathbf{n}}_B)\sin \gamma_A \sin \gamma_B \sin \zeta_A \sin \zeta_B \end{aligned} \quad \text{(G37)}$$

However, unlike in the first derivation, the six basis vectors ($\hat{\mathbf{s}}_A, \hat{\mathbf{m}}_A, \hat{\mathbf{n}}_A, \hat{\mathbf{s}}_A, \hat{\mathbf{m}}_A, \hat{\mathbf{n}}_A$) are not known explicitly. Nevertheless, the dot products can be calculated from the angular constraints

$$\hat{\mathbf{s}}_A \cdot \hat{\mathbf{s}}_B = \cos\theta_2 \tag{G38}$$

$$\hat{\mathbf{s}}_A \cdot \hat{\mathbf{m}}_B = \left(\frac{1}{\sin\delta_B}\right)[\cos\delta_A - \cos\theta_2 \cos\delta_B] \tag{G39}$$

$$\hat{\mathbf{s}}_A \cdot \hat{\mathbf{n}}_B = \left(\frac{1}{\sin\delta_B}\right)\sin\epsilon_A \sin\epsilon_B \sin\Omega_3 \tag{G40}$$

$$\hat{\mathbf{m}}_A \cdot \hat{\mathbf{s}}_B = \left(\frac{1}{\sin\delta_A}\right)[\cos\delta_B - \cos\theta_2 \cos\delta_A] \tag{G41}$$

$$\hat{\mathbf{m}}_A \cdot \hat{\mathbf{m}}_B = \left(\frac{1}{\sin\delta_A}\right)\left(\frac{1}{\sin\delta_B}\right) \times \left[1 - (\cos\delta_A)^2 - (\cos\delta_B)^2 - \cos\delta_A \cos\delta_B \cos\theta_2\right] \tag{G42}$$

$$\hat{\mathbf{m}}_A \cdot \hat{\mathbf{n}}_B = \left(\frac{-1}{\sin\delta_A}\right)\left(\frac{1}{\sin\delta_B}\right) \times \cos\delta_A \sin\epsilon_A \sin\epsilon_B \sin\Omega_3 \tag{G43}$$

$$\hat{\mathbf{n}}_A \cdot \hat{\mathbf{s}}_B = \left(\frac{-1}{\sin\delta_A}\right)\sin\epsilon_A \sin\epsilon_B \sin\Omega_3 \tag{G44}$$

$$\hat{\mathbf{n}}_A \cdot \hat{\mathbf{m}}_B = \left(\frac{1}{\sin\delta_A}\right)\left(\frac{1}{\sin\delta_B}\right) \cos\delta_B \sin\epsilon_A \sin\epsilon_B \sin\Omega_3 \tag{G45}$$

$$\hat{\mathbf{n}}_A \cdot \hat{\mathbf{n}}_B = \left(\frac{1}{\sin\delta_A}\right)\left(\frac{1}{\sin\delta_B}\right) [\cos\theta_2 - \cos\delta_A \cos\delta_B] \tag{G46}$$

Inserting these results into eq. (G37) and thence into eq. (G26) yields an equation relating ζ_A and ζ_B

$$G_0 + G_1\cos\zeta_A + G_2\sin\zeta_A + G_3\cos\zeta_B + G_4\sin\zeta_B + G_5\cos\zeta_A\cos\zeta_B + G_6\cos\zeta_A\sin\zeta_B + G_7\sin\zeta_A\cos\zeta_B + G_8\sin\zeta_A\sin\zeta_B = 0 \tag{G47}$$

where the coefficients G_0–G_8 are defined in terms of the fixed angles (γ_A, γ_B, δ_A, δ_B, ϵ_A, ϵ_B, θ_2, and Ω_3) and fixed distances (r_1, r_2, q_A, and q_B) of the problem.

Using the trigonometric half-angle identities (G3) and (G4), this equation may be converted into a second biquadratic equation in the variables z_A and z_B

$$H_0 + H_1 z_A + H_2 z_B + H_3 z_A z_B + H_4 z_A^2 + H_5 z_B^2 + H_6 z_A^2 z_B + H_7 z_A z_B^2 + H_8 z_A^2 z_B^2 = 0 \tag{G48}$$

where the nine coefficients H_0–H_8 are defined by the relations

$$H_0 \equiv G_0 + G_1 + G_3 + G_5 \tag{G49}$$

$$H_1 \equiv 2(G_2 + G_7) \tag{G50}$$

$$H_2 \equiv 2(G_4 + G_6) \tag{G51}$$

$$H_3 \equiv 4G_8 \tag{G52}$$

$$H_4 \equiv G_0 - G_1 + G_3 - G_5 \tag{G53}$$

$$H_5 \equiv G_0 + G_1 - G_3 - G_5 \tag{G54}$$

$$H_6 \equiv 2(G_4 - G_6) \tag{G55}$$

$$H_7 \equiv 2(G_2 - G_7) \tag{G56}$$

$$H_8 \equiv G_0 - G_1 - G_3 + G_5 \tag{G57}$$

This is the second biquadratic equation for eight-atom loops.

THE THIRD EQUATION

The third and fourth equations are relatively easy to derive. The vector equation (G5) for the central quadrangle can also be written

$$\mathbf{r}_1 - \mathbf{q}_A = \mathbf{r}_2 - \mathbf{q}_B \tag{G58}$$

which, when squared, gives the equation

$$r_1^2 + q_A^2 - 2r_1 q_A \cos\xi_A = r_2^2 + q_B^2 - 2r_2 q_B \cos\eta_B \tag{G59}$$

where the angles ξ_A and η_B are defined as in Figure 15. The law of cosines for spherical triangles allows these angles to be expressed in terms of Ω_A and ζ_B, respectively

$$\cos\xi_A = \cos\alpha_A \cos\beta_A + \sin\alpha_A \sin\beta_A \cos\Omega_A \tag{G60}$$

$$\cos\eta_B = \cos\gamma_B \cos\delta_B + \sin\gamma_B \sin\delta_B \cos\zeta_B \tag{G61}$$

Substituting these equations into eq. (G59) yields a relation between the spherical angles Ω_A and ζ_B.

$$I_0 + I_1\cos\Omega_A + I_2\cos\zeta_B = 0 \tag{G62}$$

Using the trigonometric half-angle identities (G1) and (G4), this relation can be converted into a biquadratic equation in the variables w_A and z_B

$$J_0 + J_1 w_A^2 + J_2 z_B^2 + J_3 w_A^2 z_B^2 = 0 \qquad \text{(G63)}$$

where the four coefficients J_0–J_3 are defined by

$$J_0 \equiv I_0 + I_1 + I_2 \qquad \text{(G64)}$$

$$J_1 \equiv I_0 - I_1 + I_2 \qquad \text{(G65)}$$

$$J_2 \equiv I_0 + I_1 - I_2 \qquad \text{(G66)}$$

$$J_3 \equiv I_0 - I_1 - I_2 \qquad \text{(G67)}$$

This is the third biquadratic equation for eight-atom loops.

THE FOURTH EQUATION

The vector equation (G5) for the central quadrangle can also be written

$$\mathbf{r}_1 + \mathbf{q}_B = \mathbf{r}_2 + \mathbf{q}_A \qquad \text{(G68)}$$

which, when squared, gives the equation

$$r_1^2 + q_B^2 - 2 r_2 q_B \cos \xi_B = r_2^2 + q_A^2 + 2 r_2 q_A \cos \eta_A \qquad \text{(G69)}$$

where the angles ξ_B and η_A are defined as in Figure 15. The law of cosines for spherical triangles allows these angles to be expressed in terms of Ω_B and ζ_A, respectively

$$\cos \xi_B = \cos \alpha_B \cos \beta_B + \sin \alpha_B \sin \beta_B \cos \Omega_B \qquad \text{(G70)}$$

$$\cos \eta_A = \cos \gamma_A \cos \delta_A + \sin \gamma_A \sin \delta_A \cos \zeta_A \qquad \text{(G71)}$$

Substituting these equations into eq. (G69) yields a relation between the spherical angles Ω_B and ζ_A.

$$X_0 + X_1 \cos \Omega_B + X_2 \cos \zeta_A = 0 \qquad \text{(G72)}$$

Using the trigonometric half-angle identities (G2) and (G3), this relation can be converted into a biquadratic equation in the variables w_B and z_A

$$Y_0 + Y_1 w_B^2 + Y_2 z_A^2 + Y_3 w_B^2 z_A^2 = 0 \qquad \text{(G73)}$$

where the four coefficients Y_0–Y_3 are defined by

$$Y_0 \equiv X_0 + X_1 + X_2 \qquad \text{(G74)}$$

$$Y_1 \equiv X_0 - X_1 + X_2 \qquad \text{(G75)}$$

$$Y_2 \equiv X_0 + X_1 - X_2 \qquad \text{(G76)}$$

$$Y_3 \equiv X_0 - X_1 - X_2 \qquad \text{(G77)}$$

This is the fourth biquadratic equation for eight-atom loops.

SOLVING THE FOUR BIQUADRATIC EQUATIONS

The valid loop closures are those solutions for w_A, w_B, z_A, and z_B, which satisfy the four defining equations (G15), (G48), (G63), and (G73). These solutions can again be found by the method of resultants.[48] Proceeding as in Appendix D, the two biquadratic equations (G15) and (G73) can be combined to eliminate the common variable w_B, yielding a biquartic equation in w_A and z_A. Similarly, eqs. (G48) and (G63) can be combined to eliminate the common variable z_B, yielding another biquartic equation in w_A and z_A. Once w_A and z_A have been solved, the remaining variables w_B and z_B have unique solutions, analogous to the solution (D25) for z_1 given in Appendix D.

Finally, these two biquartic equations can be combined to eliminate the variable z_A and obtain a polynomial equation for w_A, again by the method of resultants. If the two equations have the general form

$$S_0 + S_1 z_A + S_2 z_A^2 + S_3 z_A^3 + S_4 z_A^4 = 0 \qquad \text{(G78)}$$

$$T_0 + T_1 z_A + T_2 z_A^2 + T_3 z_A^3 + T_4 z_A^4 = 0 \qquad \text{(G79)}$$

where the coefficients S_0–S_4 and T_0–T_4 are polynomials in w_A, the resultant in w_A equals the determinant of the 8×8 square matrix

$$\begin{bmatrix} S_0 & S_1 & S_2 & S_3 & S_4 & 0 & 0 & 0 \\ 0 & S_0 & S_1 & S_2 & S_3 & S_4 & 0 & 0 \\ 0 & 0 & S_0 & S_1 & S_2 & S_3 & S_4 & 0 \\ 0 & 0 & 0 & S_0 & S_1 & S_2 & S_3 & S_4 \\ T_0 & T_1 & T_2 & T_3 & T_4 & 0 & 0 & 0 \\ 0 & T_0 & T_1 & T_2 & T_3 & T_4 & 0 & 0 \\ 0 & 0 & T_0 & T_1 & T_2 & T_3 & T_4 & 0 \\ 0 & 0 & 0 & T_0 & T_1 & T_2 & T_3 & T_4 \end{bmatrix} \qquad \text{(G80)}$$

Because the elements of this matrix are generally polynomials of degree four in w_A, its determinant (the resultant) is a polynomial of degree 32 in w_A.

The theory of resultants indicates that the two biquartic equations [(G78) and (G79)] in w_A and z_A have a common solution if and only if w_A is a root of this 32nd degree polynomial. Unlike the seven-atom and nine-atom loops treated in Appendix D, however, it is not immediately obvious how to

solve for the other variable z_A given a particular root w_A. By combining the two biquartic equations [(G78) and (G79)], the quartic term in z_A can be eliminated; but this still leaves a cubic equation in z_A, which has three possible roots, any one of which could conceivably be a valid solution of both biquartic equations. Therefore, the eight-atom loop has at most $3 \times 32 = 96$ solutions.

This limit may be restricted still further. Bezout's Theorem states that two (nondegenerate) polynomial equations in two variables of degrees d_1 and d_2 have at most $d_1 d_2$ real points of intersection, counting multiplicities, complex intersections and intersections at infinity.[50] The degree of the two biquartic equations [(G78) and (G79)] is eight, so there can be (at most) $8 \times 8 = 64$ real intersections.

References

1. Gō, N.; Scheraga, H. A. Macromolecules 1970, 3, 178.
2. Gō, N.; Scheraga, H. A. Macromolecules 1973, 6, 273.
3. Scheraga, H. A. In Large Ring Molecules; Semlyen, J. A., Ed.; John Wiley & Sons: New York, 1996, p. 99.
4. Dygert, M.; Gō, N.; Scheraga, H. A. Macromolecules 1975, 8, 750.
5. Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. Numerical Recipes in C; Cambridge University Press: New York, 1992, 2nd ed.
6. Gō, N.; Scheraga, H. A. Macromolecules 1970, 3, 188.
7. Palmer, K. A.; Scheraga, H. A. J Comput Chem 1991, 12, 505.
8. Palmer, K. A.; Scheraga, H. A. J Comput Chem 1992, 13, 329.
9. Bruccoleri, R. E.; Karplus, M. Macromolecules 1985, 18, 2767.
10. Gō, N.; Scheraga, H. A. Macromolecules 1973, 6, 525.
11. Gō, N.; Scheraga, H. A. Macromolecules 1978, 11, 552.
12. Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. J Phys Chem 1975, 79, 2361.
13. Shenkin, P. S.; Yarmush, D. L.; Fine, R. M.; Wang, H.; Levinthal, C. Biopolymers 1987, 26, 2053.
14. Liwo, A.; Tempczyk, A.; Grzonka, Z. J Comput Aided Mol Des 1989, 2, 281.
15. Zheng, Q.; Rosenfeld, R.; Vajda, S.; DeLisi, C. J Comput Chem 1993, 14, 556.
16. Braun, W. Biopolymers 1987, 26, 1691.
17. Abagyan, R. A.; Mazur, A. K. J Biomol Struct Dynam 1989, 6, 833.
18. Gibson, K. D.; Scheraga, H. A. J Comput Chem 1997, 18, 403.
19. Havel, T. F.; Kuntz, I. D.; Crippen, G. M. Bull Math Biol 1983, 45, 665.
20. Weiner, P. K.; Profeta, S., Jr.; Wipff, G.; Havel, T.; Kuntz, I. D.; Langridge, R.; Kollman, P. A. Tetrahedron 1983, 39, 1113.
21. Moult, J.; James, M. N. G. Proteins 1986, 1, 146.
22. Blundell, T. L.; Sibanda, B. L.; Sternberg, M. J. E.; Thornton, J. M. Nature 1987, 326, 347.
23. Claessens, M.; van Cutsem, E.; Lasters, I.; Wodak, S. Protein Eng 1989, 2, 335.
24. Summers, N. L.; Karplus, M. J Mol Biol 1990, 216, 991.
25. Brünger, A. T.; Clore, G. M.; Gronenborn, A. M.; Karplus, M. Proc Natl Acad Sci USA 1986, 83, 3801.
26. Braun, W. Q Rev Biophys 1987, 19, 115.
27. Jones, T. A.; Thirup, S. EMBO J 1986, 5, 819.
28. Bruccoleri, R. E.; Karplus, M. Biopolymers 1987, 26, 137.
29. Bruccoleri, R. E.; Haber, E.; Novotný, J. Nature 1988, 335, 564.
30. Greer, J. Proteins 1990, 7, 317.
31. Rosenfeld, R.; Zheng, Q.; Vajda, S.; DeLisi, C. J Mol Biol 1993, 234, 515.
32. Némethy, G.; Pottle, M. S.; Scheraga, H. A. J Phys Chem 1983, 87, 1883.
33. Némethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. J Phys Chem 1992, 96, 6472.
34. Brant, D. A.; Flory, P. J. J Am Chem Soc 1965, 87, 2791.
35. Bronshtein, I. N.; Semendyayev, K. A. Handbook of Mathematics; Van Nostrand Reinhold: New York, 1985, 3rd revised ed.
36. van der Waerden, B. L. Algebra; Springer Verlag: New York, 1991, vol I.
37. Pabo, C. O.; Suchanek, E. G. Biochemistry 1986, 25, 5987.
38. Tarnowska, M.; Liwo, A.; Shenderovich, M. D.; Liepiņa, I.; Golbraikh, A. A.; Grzonka, Z.; Tempczyk, A. J Comput Aided Mol Des 1993, 7, 699.
39. Flory, P. J. Principles of Polymer Chemistry; Cornell University Press: Ithaca, NY, 1953.
40. de Gennes, P.-G. Scaling Concepts in Polymer Physics; Cornell University Press: Ithaca, NY, 1979.
41. Alexandrowicz, Z. J Chem Phys 1967, 46, 3800.
42. Purisima, E. O.; Scheraga, H. A. Biopolymers 1984, 23, 1207.
43. Wako, H.; Scheraga, H. A. J Protein Chem 1982, 1, 5.
44. Vasmatzis, G.; Brower, R.; Delisi, C. Biopolymers 1994, 34, 1669.
45. Oberlin, D., Jr.; Scheraga, H. A. J Comput Chem 1998, 19, 71.
46. Swendsen, R. H.; Wang, J.-S. Phys Rev Lett 1987, 58, 86.
47. Lin, S. Bell Syst Tech J 1965, 44, 2245.
48. Zwillinger, D. Handbook of Differential Equations; Academic Press: San Diego, 1992, 2nd ed.
49. Char, B. W.; Geddes, K. O.; Gonnet, G. H.; Leong, B. L.; Monagan, M. B.; Watt, S. M. Maple V: Language Reference Manual; Springer Verlag: New York, 1991.
50. Borowski, E. J.; Borwein, J. M. The HarperCollins Dictionary of Mathematics; HarperCollins Publishers: New York, 1991.